

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 15:20:40 ; Search time 3001 Seconds  
(without alignments)  
13731.947 Million cell updates/sec

Title: US-09-357-675C-1  
Perfect score: 1416  
Sequence: 1 gccacactgcgtgcgcctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1307	92.3	1385	9	AF069987
2	571	40.3	4079	9	AF069984 Homo sapi
3	571	40.3	167863	2	AL590651 Homo sapi
4	571	40.3	200822	9	AL591806 Human DNA
5	326	23.0	377	6	AX368386 Sequence
6	173	12.2	441	11	G13436 human STS W
7	87	6.1	2261	6	AX397469 Sequence
8	87	6.1	2261	9	AK001497 Homo sapi
9	47	3.3	847	11	G72919 MARC 2849-2
10	33	2.3	1338	10	AF069988 Mus muscu
11	33	2.3	1365	10	BC021634 Mus muscu
12	33	2.3	4481	10	AF069985 Mus muscu
13	33	2.3	181583	2	AC105589 Rattus no
14	33	2.3	211772	10	AC084821 Mus muscu
15	33	2.3	215043	10	AC087229 Mus muscu
16	32	2.3	115667	2	AC125857 Rattus no
17	31	2.2	144986	2	AC108571 Rattus no
18	30	2.1	1587	9	BC017413 Homo sapi
19	30	2.1	2421	10	BC016425 Mus muscu
20	30	2.1	3300	6	AR212407 Sequence
21	30	2.1	4113	10	AF230376 Meriones
22	30	2.1	44645	9	AC108058 Homo sapi
23	30	2.1	60773	2	AC069576 Homo sapi
24	30	2.1	252059	2	AC102081 Mus muscu
25	29	2.0	345	10	RNRIPRL38 X57007 Rat mRNA fo
26	29	2.0	549	9	AF112344 Homo sapi
27	29	2.0	564	6	AX407609 Sequence
28	29	2.0	876	5	AF175983 Rana sylv
29	29	2.0	1049	8	AF071889 Prunus ar
30	29	2.0	1075	9	BC014385 Homo sapi
31	29	2.0	1134	10	AF058799 Mus muscu
32	29	2.0	1193	6	A29423 putative bo
33	29	2.0	1193	6	AR022483 Sequence
34	29	2.0	1193	6	AR068961 Sequence
35	29	2.0	1193	6	AR103330 Sequence
36	29	2.0	1193	6	AR134988 Sequence
37	29	2.0	1193	6	AR141817 Sequence
38	29	2.0	1193	6	AR143334 Sequence
39	29	2.0	1193	6	AR151943 Sequence
40	29	2.0	1193	6	I36423 Sequence 13
41	29	2.0	1193	6	I88123 Sequence 13
42	29	2.0	1211	5	AY065841 Danio rer
43	29	2.0	1267	10	BC004064 Mus muscu
44	29	2.0	1488	9	HSU79258 Human clone
45	29	2.0	1667	3	AY047568 Drosophila

#### ALIGNMENTS

RESULT 1  
AF069987  
LOCUS AF069987 1385 bp mRNA linear PRI 23-JUL-1998  
DEFINITION Homo sapiens nitrlase 1 (NIT1) mRNA, complete cds.  
ACCESSION AF069987  
VERSION AF069987.1 GI:3228665  
KEYWORDS SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1385)  
Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,  
Tillib, S., Draganes, A., Wernuth, P., Rothman, J.H., Huebner, K.,  
Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

TITLE Nitrilase and Fhit homologs are encoded as fusion proteins in  
JOURNAL Drosophila melanogaster and Caenorhabditis elegans  
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
PUBMED 98337986  
REFERENCE 2 (bases 1 to 1385)  
AUTHORS Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,  
Tillib, S., Draganescu, A., Wermuth, P., Rothman, J., Huebner, K.,  
Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.  
Direct Submission  
Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA  
FEATURES  
Location/Qualifiers  
1..1385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21-q22"  
1..1385  
/gene="NIT1"  
77..1060  
/gene="NIT1"  
/note="NIT1"  
/codon\_start=1  
/product="nitrilase 1"  
/protein\_id="AAC39907.1"  
/db\_xref="GI:3228666"  
/translation="MLGFTIRPPHRLSLPGLRIPOLSVLCQAPRRMAISSSS  
ELPLVAVCVSTPDQKNFXTCAELVREARGLAFLEAFEDFTARDPAETLHLS  
EPLGKLEETOLARECGLWLSGGFHERQDWEOQKIYNCHVLNLSKGVATYR  
KHLCDVEIPQGGPMSNSMPGSLSPVSTPAGKIGLAVCYDMRFPPELSLALQA  
GAEILTVPSAFSGITGPAHVEVLLRRAIETQCVVAAACGRHHEKRAVSGHSMVVD  
PWGTVARCSGPGGLCLARIDLNLRLRRLPVFQHRPDLNGLHPLS"  
BASE COUNT 322 a 382 c 357 g 324 t  
ORIGIN  
Query Match 92.3%; Score 1307; DB 9; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 110 GCTGGGCTTCATCACCAGGCCCTCCTCACAGATTCCTGTCCTTCTGTGTCCTGGACTCGG 169  
DB 79 GCTGGGCTTCATCACCAGGCCCTCCTCACAGATTCCTGTCCTTCTGTGTCCTGGACTCGG 138  
QY 170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCAGGCGCATGGCTATCTCCTC 229  
DB 139 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCAGGCGCATGGCTATCTCCTC 198  
QY 230 TTCTCCTCGCAACTGCCCTGGTGGCTGTGTGCCAGGTAAACATCGAGCCAGACAAGCA 289  
DB 199 TTCTCCTCGCAACTGCCCTGGTGGCTGTGTGCCAGGTAAACATCGAGCCAGACAAGCA 258  
QY 290 ACAGAACTTTAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCT 349  
DB 259 ACAGAACTTTAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCT 318  
QY 350 GCCTTCTCTGCTGAGGATTTGACTTTCATTCAGCGGACCTCGCAGAGCGCTACACCT 409  
DB 319 GCCTTCTCTGCTGAGGATTTGACTTTCATTCAGCGGACCTCGCAGAGCGCTACACCT 378  
QY 410 GTCTGAACCACTGGTGGGAACTTTTGAAGAATACACCCAGCTTGGCAGGAAATGGG 469  
DB 379 GTCTGAACCACTGGTGGGAACTTTTGAAGAATACACCCAGCTTGGCAGGAAATGGG 438  
QY 470 ACTCTGGCTGCTTGGGTGGTTCCTCAGGCTGGCGGCAAGACTGGGAGCACTCAGAA 529  
DB 439 ACTCTGGCTGCTTGGGTGGTTCCTCAGGCTGGCGGCAAGACTGGGAGCACTCAGAA 498  
QY 530 AATCTACAATTTGCTACGTGCTGCTGAACAGCAAGAGGCGAGTAGTGGCCACTTACAGAA 589  
DB 499 AATCTACAATTTGCTACGTGCTGCTGAACAGCAAGAGGCGAGTAGTGGCCACTTACAGAA 558  
QY 590 GACACATCTGTGTGACGTAGAGATTTCCAGGGCAGGGGCTATGTGTGAAGCAACTCTAC 649

Db 559 GACACATCTGTGTGACGTAGAGATTCAGGGCAGGGGCTATGTGTGAAGCAACTCTAC 618  
QY 650 CATGGCTGGGCCAGCTCTTGAGTCACTGTGAGCAGACACAGCAGGCAAGATGGTCTAGC 709  
Db 619 CATGGCTGGGCCAGCTCTTGAGTCACTGTGAGCAGACACAGCAGGCAAGATGGTCTAGC 678  
QY 710 TGTCTGCTATGACATGCGGTTCCTGAACTCTCTCTGCACTTGGCTCAAGCTGGAGCAGA 769  
Db 679 TGTCTGCTATGACATGCGGTTCCTGAACTCTCTCTGCACTTGGCTCAAGCTGGAGCAGA 738  
QY 770 GATACCTTACCTATCTCTCAGCTTTTGGATTCATTACAGGCCAGGCCACTGGGAGGTGT 829  
Db 739 GATACCTTACCTATCTCTCAGCTTTTGGATTCATTACAGGCCAGGCCACTGGGAGGTGT 798  
QY 830 GCTGGGGCCGCTGCTATCTGAAACCCAGTGTATGTAGTGGCAGCAGCAGCTGTGGAGC 889  
Db 799 GCTGGGGCCGCTGCTATCTGAAACCCAGTGTATGTAGTGGCAGCAGCAGCTGTGGAGC 858  
QY 890 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGTGTAGACCCCTGGGAAACAGT 949  
Db 859 CCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGTGTAGACCCCTGGGAAACAGT 918  
QY 950 GGTGGCCCGCTGCTCTGAGGGGCCAGGGCTCTGCTTGGCCGAATAGACCTCAACTATCT 1009  
Db 919 GGTGGCCCGCTGCTCTGAGGGGCCAGGGCTCTGCTTGGCCGAATAGACCTCAACTATCT 978  
QY 1010 GCGACAGTTGGGCCGACACCTGCTGTGTCAGACACGCGAGGCTGACCTCTATGGCAA 1069  
Db 979 GCGACAGTTGGGCCGACACCTGCTGTGTCAGACACGCGAGGCTGACCTCTATGGCAA 1038  
QY 1070 TCTGGGTCAACCACTGTCTTAAGACTTGTCTGTAGTTTGTAGACCTGCCCTCCACCC 1129  
Db 1039 TCTGGGTCAACCACTGTCTTAAGACTTGTCTGTAGTTTGTAGACCTGCCCTCCACCC 1098  
QY 1130 CCCACCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGGCACAGCT 1189  
Db 1099 CCCACCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGGCACAGCT 1158  
QY 1190 CCCCTCACTTGGAGAACCTTGAATCTCTTGTATGGAAACAGATGGGCTGTGGGAAAGA 1249  
Db 1159 CCCCTCACTTGGAGAACCTTGAATCTCTTGTATGGAAACAGATGGGCTGTGGGAAAGA 1218  
QY 1250 AACTTTCACCTGAGCTTTCACCTGAGGTGAGTGTGAGTGTGGAGAGGTGGAAATTTATA 1309  
Db 1219 AACTTTCACCTGAGCTTTCACCTGAGGTGAGTGTGAGTGTGGAGAGGTGGAAATTTATA 1278  
QY 1310 TAGTCATTTTATTTATGAGAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1369  
Db 1279 TAGTCATTTTATTTATGAGAACTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGCATT 1338  
QY 1370 GAAAT 1416  
Db 1339 GAAAT 1385  
RESULT 2  
LOCUS AF069984 4079 bp DNA linear PRI 23-JUL-1998  
DEFINITION Homo sapiens nitrilase homolog 1 (NIT1) gene, alternatively spliced  
product, complete cds.  
ACCESSION AF069984  
VERSION AF069984.1 GI:3242977  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 4079)  
REFERENCE  
AUTHORS Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,  
Tillib, S., Draganescu, A., Wermuth, P., Rothman, J.H., Huebner, K.,  
Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.  
Nitrilase and Fhit homologs are encoded as fusion proteins in

JOURNAL Drosophila melanogaster and Caenorhabditis elegans  
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
98337986  
PUBMED 9671749  
REFERENCE 2 (bases 1 to 4079)  
AUTHORS Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,  
Tillib, S., Draganes, A., Wermuth, P., Rothman, J., Huebner, K.,  
Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA  
FEATURES  
Source Location/Qualifiers  
1..4079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21-q22"  
635..3569  
/gene="NIT1"  
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3569)  
/gene="NIT1"  
/product="nitric oxide synthase homolog 1"  
/note="alternatively spliced"  
635..680  
/gene="NIT1"  
/note="alternatively spliced"  
/number=1  
join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3267)  
/gene="NIT1"  
/note="NIT1; alternatively spliced"  
/codon\_start=1  
/product="nitric oxide synthase homolog 1"  
/protein\_id="AAC39901.1"  
/db\_xref="GI:3242978"  
/translation="MLGFTTRPPHRLSLCPGIRIPOLSVLCQPRPRAWISSSS  
ELPLVAVCTSTDPKQNFKAELVREARLACLAFLPEAFDFIARDPAETLHLS  
EPLGKLEIYETQLAREGFWLGLGPHERGQDWEOTQKIYCHVLNLSKGVAVTVR  
KTHLDCVEIPQSGPMCSNTPGSPLESVPSTGKIGLAVCYDMRPELSLALQA  
GAETLTVPSAGSITGPAHVEVLLRAETQCYVAAACGGRHHEKRSYGHSMYVD  
PWGTVARCSEGPGLCLARDLNYLRRLPVPFQRRPDLNGLGHLPS"  
681..711  
/gene="NIT1"  
/note="1A; alternatively spliced"  
712..741  
/gene="NIT1"  
/note="1B; alternatively spliced"  
940..1053  
/gene="NIT1"  
/note="1C; alternatively spliced"  
1288..1383  
/gene="NIT1"  
/number=2  
1636..1890  
/gene="NIT1"  
/number=3  
2011..2114  
/gene="NIT1"  
/number=4  
2321..2454  
/gene="NIT1"  
/number=5  
2609..2734  
/gene="NIT1"  
/number=6  
3001..3569  
/gene="NIT1"  
/number=7  
940 a 1082 c 1043 g 1012 t 2 others

Query Match

40.3%; Score 571; DB 9; Length 4079;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 823 AGGTGTTCTGCGGGCCGCTGCTATCGAAACCCAGTGTATAGTGGCAGCAGCAGT 882  
|||||  
DB 2999 AGGTGTTCTGCGGGCCGCTGCTATCGAAACCCAGTGTATAGTGGCAGCAGCAGT 3058  
|||||  
QY 883 GTGAGCGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGTGTAGACCCCTGGG 942  
|||||  
DB 3059 GTGAGCGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGTGTAGACCCCTGGG 3118  
|||||  
QY 943 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCCTCTGCCCTCCCGCAATAGACCTCA 1002  
|||||  
DB 3119 GAACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCCCTCTGCCCTCCCGCAATAGACCTCA 3178  
|||||  
QY 1003 ACTATCTGCGACAGTTGGCGCGACACCTGCCCTGTTCCAGCACCGCAGGCCCTGACCTCT 1062  
|||||  
DB 3179 ACTATCTGCGACAGTTGGCGCGACACCTGCCCTGTTCCAGCACCGCAGGCCCTGACCTCT 3238  
|||||  
QY 1063 ATGCAATCTGGGTGTCACCCACTGTCTTAAGACTTGACTTCTGTAGTTTAGACCTGCCCC 1122  
|||||  
DB 3239 ATGCAATCTGGGTGTCACCCACTGTCTTAAGACTTGACTTCTGTAGTTTAGACCTGCCCC 3298  
|||||  
QY 1123 TCCACCCCGCCCTGCGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGG 1182  
|||||  
DB 3299 TCCACCCCGCCCTGCGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGG 3358  
|||||  
QY 1183 CACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGTATGGAACACAGATGGCTGCTTG 1242  
|||||  
DB 3359 CACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGTATGGAACACAGATGGCTGCTTG 3418  
|||||  
QY 1243 GGAAGAAACTTTCACCTGAGCTTCACCTGAGGTGAGTGCAGTTCAGAAAGGTGGAA 1302  
|||||  
DB 3419 GGAAGAAACTTTCACCTGAGCTTCACCTGAGGTGAGTGCAGTTCAGAAAGGTGGAA 3478  
|||||  
QY 1303 TTTTATAGTCATGTTTATTTATTCATGGAACCTGAAGTTCCTCTGAGGCTGAGCAGCAC 1362  
|||||  
DB 3479 TTTTATAGTCATGTTTATTTATTCATGGAACCTGAAGTTCCTCTGAGGCTGAGCAGCAC 3538  
|||||  
QY 1363 TGGCATTGAAAATAATAATATCATAAAGTC 1393  
DB 3539 TGGCATTGAAAATAATAATATCATAAAGTC 3569  
RESULT 3  
AL590651/c  
LOCUS Homo sapiens chromosome 1 clone RP11-137A12, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 9 unordered pieces.  
ACCESSION AL590651  
VERSION AL590651.4 GI:13990236  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 167863)  
AUTHORS Harrison, E.  
TITLE Direct Submission  
JOURNAL Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On May 7, 2001 this sequence version replaced gi:13751001.  
COMMENT  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: bal37A12  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 164729 bases at least Q40  
 Consensus quality: 165486 bases at least Q30  
 Consensus quality: 166169 bases at least Q20  
 Insert size: 167063; sum-of-contigs  
 Quality coverage: 8.28x in Q20 bases; agarose-fp  
 coverage: 8.11x in Q20 bases; agarose-fp

\*\*\*\*\*  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 \* 9373: contig of 9373 bp in length  
 \* 9374 9473: gap of 100 bp  
 \* 9474 29891: contig of 20418 bp in length  
 \* 29892 29991: gap of 100 bp  
 \* 29992 36239: contig of 6248 bp in length  
 \* 36240 36339: gap of 100 bp  
 \* 36340 103784: contig of 67445 bp in length  
 \* 103785 103884: gap of 100 bp  
 \* 103885 114878: contig of 10994 bp in length  
 \* 114879 114978: gap of 100 bp  
 \* 114979 126347: contig of 11369 bp in length  
 \* 126348 126447: gap of 100 bp  
 \* 126448 137764: contig of 11317 bp in length  
 \* 137765 137864: gap of 100 bp  
 \* 137865 163160: contig of 25296 bp in length  
 \* 163161 163260: gap of 100 bp  
 \* 163261 167863: contig of 4603 bp in length.

#### FEATURES

source  
 1. 167863  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-137A12"  
 /clone\_lib="RPC1-11.1"  
 1. 9373  
 /note="assembly\_fragment:02360  
 fragment\_chain:1  
 clone\_end:T7  
 vector\_side:left"  
 9474. 29891  
 /note="assembly\_fragment:01854  
 fragment\_chain:1"  
 29992. 36239  
 /note="assembly\_fragment:00805  
 fragment\_chain:1"  
 36340. 103784  
 /note="assembly\_fragment:01467  
 fragment\_chain:1"  
 103885. 114878  
 /note="assembly\_fragment:00991  
 fragment\_chain:2"  
 114979. 126347  
 /note="assembly\_fragment:01503  
 fragment\_chain:2"  
 126448. 137764  
 /note="assembly\_fragment:01297  
 fragment\_chain:2"  
 137865. 163160  
 /note="assembly\_fragment:00117  
 fragment\_chain:2"  
 163261. 167863  
 /note="assembly\_fragment:01267  
 fragment\_chain:2  
 clone\_end:SP6  
 vector\_side:right"  
 BASE COUNT 43927 a 39935 c 40002 g 43197 t 802 others

#### ORIGIN

Query Match 40.3%; Score 571; DB 2; Length 167863;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 823 AGGTGTTGCTCGGCGCGTCTATCGAAACCCAGTGTATGTAGTGGCAGCAGCAGT 882  
 |||||  
 Db 149682 AGGTGTTGCTCGGCGCGTCTATCGNACCAGTGTATGTAGTGGCAGCAGCAGT 149623  
 |||||  
 QY 883 GTGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGTAGACCCCTGGG 942  
 |||||  
 Db 149622 GTGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGGTGTAGACCCCTGGG 149563  
 |||||  
 QY 943 CAACAGTGGTGGCGCGTCTCTGAGGGGCCAGGCTTGCCTTGCCTGCGCGAATAGACCTCA 1002  
 |||||  
 Db 149562 GAACAGTGGTGGCGCGTCTCTGAGGGGCCAGGCTTGCCTTGCCTGCGCGAATAGACCTCA 149503  
 |||||  
 QY 1003 ACTATCTGCGACAGTTGCGCGGACACCTGCCTGTGTTCCAGCACCGCAGGCTTGACCTCT 1062  
 |||||  
 Db 149502 ACTATCTGCGACAGTTGCGCGGACACCTGCCTGTGTTCCAGCACCGCAGGCTTGACCTCT 149443  
 |||||  
 QY 1063 ATGCAATCTGGGTGACCCACTGCTTTAAGACTTGACTTCTGTGAGTTAGACCTGCCCC 1122  
 |||||  
 Db 149442 ATGCAATCTGGGTGACCCACTGCTTTAAGACTTGACTTCTGTGAGTTAGACCTGCCCC 149383  
 |||||  
 QY 1123 TCCACACCCACCCCTGCCACTATGAGCTAGTGTCTGCTGACTTGGAGGAGGATCCAGG 1182  
 |||||  
 Db 149382 TCCACACCCACCCCTGCCACTATGAGCTAGTGTCTGCTGACTTGGAGGAGGATCCAGG 149323  
 |||||  
 QY 1183 CACAGCTCCCTCTACTTGGGAACTTGACCTCTGATGGAACACAGATGGCTGCTTG 1242  
 |||||  
 Db 149322 CACAGCTCCCTCTACTTGGGAACTTGACCTCTGATGGAACACAGATGGCTGCTTG 149263  
 |||||  
 QY 1243 GGAAGAACTTTCCACTGAGCTTCCACTGAGGTGACACTGCACTTTTCAGAAAGGTGAA 1302  
 |||||  
 Db 149262 GGAAGAACTTTCCACTGAGCTTCCACTGAGGTGACACTGCACTTTTCAGAAAGGTGAA 149203  
 |||||  
 QY 1303 TTTTATATAGTCATGTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGGCTGAGCAGCAC 1362  
 |||||  
 Db 149202 TTTTATATAGTCATGTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGGCTGAGCAGCAC 149143  
 |||||  
 QY 1363 TGGCATTGAAAAATATAATATCAATAAGTC 1393  
 |||||  
 Db 149142 TGGCATTGAAAAATATAATATCAATAAGTC 149112

#### RESULT 4

AL591806 Human DNA sequence from clone RP11-544M22 on chromosome 1, complete  
 LOCUS sequence.  
 DEFINITION

AL591806  
 AL591806.16 GI:18476709  
 HTG.

#### ORGANISM

human.

#### REFERENCE

1 (bases 1 to 200822)  
 Direct Submission  
 Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Feb 1, 2002 this sequence version replaced gi:17902927.

#### AUTHORS

Harrison, E.

#### TITLE

JOURNAL

#### COMMENT

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all



regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep). This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>. Rpl1-544M22 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

This sequence is the entire insert of clone RPl1-544M22. The true left end of clone RPl1-137A12 is at 156538 in this sequence. The true right end of clone RPl1-381D2 is at 145015 in this sequence.

## FEATURES

Source

1..200822  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RPl1-544M22"  
/clone\_lib="RPl1-11.2"  
10395..10397

misc\_feature

/note="Sequence from overlapping clone RPl1-381D2 (AL162592). Assembly confirmed by restriction digest."  
44098..44107

misc\_feature

/note="Sequence from overlapping clone RPl1-381D2 (AL162592). Assembly confirmed by restriction digest."  
59585..59604

misc\_feature

/note="Sequence from overlapping clone RPl1-381D2 (AL162592). Assembly confirmed by restriction digest."  
90797..90883

misc\_feature

/note="Single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."  
90861..90883

misc\_feature

/note="Sequence from uni-directional dGTP big dye terminator reads only."

BASE COUNT 55215 a 49052 c 46183 g 50372 t

ORIGIN

Query Match 40.3%; Score 571; DB 9; Length 200822;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 823 AGGTGTTGTCGGGGCCGTCATCGAACCACGAGTCATGTCAGTGGCAGCAGCAGT 882  
|||||  
Db 174499 AGGTGTTGTCGGGGCCGTCATCGAACCACGAGTCATGTCAGTGGCAGCAGCAGT 174558  
|||||

QY 883 GTGGACCCACCATTGAGAGAGAGCAAGTATGCGCCACACGATGGTGGTAGACCCCTGGG 942  
|||||  
Db 174559 GTGGACCCACCATTGAGAGAGAGCAAGTATGCGCCACACGATGGTGGTAGACCCCTGGG 174618  
|||||

QY 943 GAACAGTGGTGGCCGCTGCTGTGAGGGGCCAGGCCCTGTCGCTTGGCCGAAATAGACCTCA 1002  
|||||  
Db 174619 GAACAGTGGTGGCCGCTGCTGTGAGGGGCCAGGCCCTGTCGCTTGGCCGAAATAGACCTCA 174678  
|||||

QY 1003 ACTATCTGCGACAGTTGGCCGACACCTGCTGTCGTCAGACACCGACGCTGACCTCT 1062  
|||||  
Db 174679 ACTATCTGCGACAGTTGGCCGACACCTGCTGTCGTCAGACACCGACGCTGACCTCT 174738  
|||||

QY 1063 ATGGCAATCTGGGTCAACCCACTGCTTTAAGACTTGAATCTGTCGAGTTTAGACCTCGCCCC 1122  
|||||  
Db 174739 ATGGCAATCTGGGTCAACCCACTGCTTTAAGACTTGAATCTGTCGAGTTTAGACCTCGCCCC 174798  
|||||

QY 1123 TCCACCCCCACCCCTGCCACTATGAGCTAGTGCATGTAAGTGGAGGAGGATCCAGG 1182  
|||||

|||||  
Db 174799 TCCACCCCCACCCCTGCCACTATGAGCTAGTGCATGTAAGTGGAGGATCCAGG 174858  
|||||

QY 1183 CACAGCTCCCTCAGCTTGGAGAACCTTGACTCTCTTGTATGGAAACAGATGGGCTGCTTG 1242  
|||||

Db 174859 CACAGCTCCCTCAGCTTGGAGAACCTTGACTCTCTTGTATGGAAACAGATGGGCTGCTTG 174918  
|||||

QY 1243 GGAAGAAACTTTCACCTCAGCTTACCTTGAGTGCAGCTGAGTTCAGAAAGTGGAA 1302  
|||||

Db 174919 GGAAGAAACTTTCACCTCAGCTTACCTTGAGTGCAGCTGAGTTCAGAAAGTGGAA 174978  
|||||

QY 1303 TTTTATATAGTCATGTTTATTTATGGAACCTGAAGTTCCTGCTGAGGCTGAGCAGCAC 1362  
|||||

Db 174979 TTTTATATAGTCATGTTTATTTATGGAACCTGAAGTTCCTGCTGAGGCTGAGCAGCAC 175038  
|||||

QY 1363 TGGCATTTGAAAATAATAATAATCATAAAGTC 1393  
|||||

Db 175039 TGGCATTTGAAAATAATAATAATCATAAAGTC 175069  
|||||

RESULT 5  
AX368386/c 377 bp DNA linear PAT 16-FEB-2002

LOCUS AX368386  
DEFINITION Sequence 1096 from Patent WO0204514.  
ACCESSION AX368386  
VERSION AX368386.1 GI:18856459  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Wang, T., Watanabe, Y., Henderson, R.A., Johnson, J.C., Retter, M.W.,  
Marnerakis, M., Carter, D., Fanger, G.R., Vedvick, T.S., Bangur, C.S.,  
McNabb, A., Fanger, N., Switzer, A., McNeill, P.D. and Clapper, J.D.  
TITILE Compositions and methods for the therapy and diagnosis of lung  
cancer  
JOURNAL Patent: WO 0204514-A 1096 17-JAN-2002;  
CORIXA CORPORATION (US)  
FEATURES  
source 1..377  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 94 a 100 c 112 g 71 t  
ORIGIN

Query Match 23.0%; Score 326; DB 6; Length 377;  
Best Local Similarity 99.7%; Pred. No. 1.7e-190;  
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 129 CTTCTTCACAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 188  
|||||

Db 377 CTTCTTCACAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318  
|||||

QY 189 CTTTGTGCTCAGCCAGCCAGAGCCATGGCTATCTCTCTTCTCTCTCTCTCTCTCTCTCTCT 248  
|||||

Db 317 CTTTGTGCTCAGCCAGCCAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 258  
|||||

QY 249 CTGGTGGCTGTGTGCCAGGTAACATCGCCAGACAGACAGACAGACATTTAAACATGT 308  
|||||

Db 257 CTGGTGGCTGTGTGCCAGGTAACATCGCCAGACAGACAGACATTTAAACATGT 198  
|||||

QY 309 GCTGAGCTGGTTCGAGAGGCTGCCAGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 368  
|||||

Db 197 GCTGGCTGGTTCGAGAGGCTGCCAGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 138  
|||||

QY 369 TTTGACTTTCATTCGACGGGACCTCGCAGAGACGCTACACCTGTCTGAAACACTGGGTGG 428  
|||||

Db 137 TTTGACTTTCATTCGACGGGACCTCGCAGAGACGCTACACCTGTCTCAACACTGGGTGG 78  
|||||

QY 429 AAACCTTTGAGAGATACACCCAGCTTGCAGGGAAATGTTGGACTCTGGCTGTCTCTTGGGT 488  
|||||

Db 77 AAACCTTTGAGAGATACACCCAGCTTGCAGGGAAATGTTGGACTCTGGCTGTCTCTTGGGT 18  
|||||

QY 489 GGTTCATGAGCGTGG 505  
 Db 17 GGTTCATGAGCGTGG 1  
 |||||||H|||||  
 RESULT 6  
 G13436/c  
 LOCUS G13436 441 bp DNA linear STS 04-JUN-1996  
 DEFINITION human STS WI-12338, sequence tagged site.  
 ACCESSION G13436  
 VERSION G13436.1 GI:1127545  
 KEYWORDS STS; STS sequence; primer; sequence tagged site.  
 SOURCE Homo sapiens STSs derived from sequences in dbEST and the Unigene collection.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 441)  
 AUTHORS Hudson,T.  
 TITLE Whitehead Institute/MIT Center for Genome Research; Physically  
 JOURNAL Mapped STSs  
 COMMENT Unpublished (1995)  
 Contact: Thomas Hudson  
 Whitehead Institute/MIT Center for Genome Research  
 Whitehead Institute for Biomedical Research  
 9 Cambridge Center, Cambridge MA 02142 USA  
 Tel: 617 252 1900  
 Fax: 617 252 1902  
 Email: thudson@genome.wi.mit.edu  
 Primer A: TATATTTTCAATGCCAGTCTG  
 Primer B: AACTTTCACCTGAGCTTCACC  
 STS size: 130  
 PCR profile:  
 Presoak:  
 Denaturation:  
 Annealing: 56 degrees C  
 Polymerization:  
 PCR Cycles: 35  
 Thermal Cycler:  
 Protocol:  
 Template: 10 ng  
 Primer: each 5 pM  
 dNTPs: each 4 nM  
 Taq Polymerase: 0.025 units/ul  
 Total Vol: 20 ul  
 Buffer:  
 MgCl2: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 9.3  
 Derived from dbEST (genbank accession T78621).  
 FEATURES  
 source  
 1..441  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="750.8 cR from top of Chr1 linkage group"  
 STS  
 primer\_bind 16..145  
 primer\_bind 16..38  
 BASE COUNT 107 a 95 c 130 g 103 t 6 others  
 ORIGIN  
 Query Match 12.2%; Score 173; DB 11; Length 441;  
 Best Local Similarity 100.0%; Pred. No. 2e-95;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1178 CCAGGCACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGCT 1237  
 |||||||H|||||

Db 216 CCAGGCACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGCT 157  
 QY 1238 GCTTGGGAAGAACTTTACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGG 1297  
 |||||||H|||||  
 Db 156 GCTTGGGAAGAACTTTACCTGAGCTTCACCTGAGGTCAGACTGCAGTTTCAGAAAGG 97  
 |||||||H|||||  
 QY 1298 TGGAAATTTATATAGTCAATGTTTATTTTCATGGAACAACTGAAGTTCTGCTGAGG 1350  
 |||||||H|||||  
 Db 96 TGGATTTTATATAGTCAATGTTTATTTTCATGGAACAACTGAAGTTCTGCTGAGG 44  
 |||||||H|||||  
 RESULT 7  
 AX397469/c  
 LOCUS AX397469 2261 bp DNA linear PAT 18-MAY-2002  
 DEFINITION Sequence 1684 from Patent WO212328.  
 ACCESSION AX397469  
 VERSION AX397469.1 GI:21068216  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS King,G.E., Meagher,M.J., Xu,J. and Secrist,H.  
 TITLE Compositions and methods for the therapy and diagnosis of colon  
 JOURNAL cancer  
 PATENT: WO 0212328-A 1684 14-FEB-2002;  
 CORIXA CORPORATION (US)  
 FEATURES  
 Location/Qualifiers  
 1..2261  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 543 a 594 c 541 g 583 t  
 ORIGIN  
 Query Match 6.1%; Score 87; DB 6; Length 2261;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-42;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1307 ATATAGTCATCTGTTTATTTTCATGGAACAACTGAAGTTCTGCTGAGGCTGAGCAGCTGGC 1366  
 |||||||H|||||  
 Db 2261 ATATAGTCATCTGTTTATTTTCATGGAACAACTGAAGTTCTGCTGAGGCTGAGCAGCTGGC 2202  
 |||||||H|||||  
 QY 1367 ATTGAAAATATAATAATCAATAAGTC 1393  
 |||||||H|||||  
 Db 2201 ATTGAAAATATAATAATCAATAAGTC 2175  
 |||||||H|||||  
 RESULT 8  
 AK001497/c  
 LOCUS AK001497 2261 bp mRNA linear PRI 01-AUG-2002  
 DEFINITION Homo sapiens cDNA FLJ10635 fis, clone NT2RP2003669, highly similar  
 to Homo sapiens mRNA for DEDD protein.  
 ACCESSION AK001497  
 VERSION AK001497.1 GI:7022790  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA,  
 clone\_lib:NT2RP2 clone:NT2RP2005669.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,  
 Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y.,  
 Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K.,  
 Masuho,Y. and Kanehori,K.  
 TITLE NEDO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2261)  
 AUTHORS Isogai,T. and Otsuki,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@nri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing and clone selection;  
Helix Research Institute (supported by Japan Key Technology Center  
etc.) and Department of Virology, Institute of Medical Science,  
University of Tokyo.

## FEATURES

Location/Qualifiers

1..2261  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP2005669"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP2"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."

BASE COUNT 543 a 594 c 541 g 583 t

ORIGIN

Query Match 6.1%; Score 87; DB 9; Length 2261;  
Best Local Similarity 100.0%; Pred. No. 7.1e-42;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 ATATAGTCATTCTTATTTTCATGGAACTGAAGTCTGCTGAGGCTGAGCAGCACTGGC 1366

Db 2261 ATATAGTCATTCTTATTTTCATGGAACTGAAGTCTGCTGAGGCTGAGCAGCACTGGC 2202

QY 1367 ATTGAAAAATAAATAATCAATAAGTC 1393

Db 2201 ATTGAAAAATAAATAATCAATAAGTC 2175

## RESULT 9

G72919

LOCUS G72919 847 bp DNA linear STS 08-AUG-2001  
DEFINITION MARC 2849-2850:991933517:1 SCF - porcine spleen Sus scrofa STS  
genomic, sequence tagged site.

ACCESSION G72919

VERSION G72919.1 GI:15146949

KEYWORDS STS.

SOURCE Sus scrofa.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 847)

AUTHORS Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L. and

Keefe, J.W.

TITLE Single nucleotide polymorphism (SNP) discovery in expressed porcine

genes

UNPUBLISHED

JOURNAL

COMMENT

Contact: Freking BA

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4278

Fax: 402 762 4173

Email: freking@email.marc.usda.gov

Primer A: GGGTTCAGCTTTGGATCTG

Primer B: GCTGAGCCTGTTTCATCA

STS size: 700

PCR Profile:

Hotstart: 95 degrees for 15 minutes

Denature: 95 degrees for 30 seconds

Anneal: 58 degrees

Extension: 68 degrees for 2 minutes

Cycles: 32 to 45

Protocol:

Template: 50-200 ng genomic DNA

Primer: each 20 pmoles

dNTPs: each 88 uM

Taq Polymerase: 0.25 units (Qiagen HotStar)

## Buffer:

Commercially supplied Qiagen HotStar buffer

The STS is derived from PCR amplicons generated from genomic DNA,  
sequenced from each end using the amplification primers. The  
sequence does not necessarily represent the entire amplicon.  
Sequence derived from PolyPhred was trimmed from each end of each  
unique contig until five consecutive bases exceeded a quality score  
threshold of 20, and the next 10 bases averaged a quality score of  
20 or greater. Amplicon size was estimated by agarose gel  
electrophoresis.

## FEATURES

source

Location/Qualifiers

1..847  
/organism="Sus scrofa"  
/strain="white composite, duroc, meishan, minzhu,  
fengling, crossbreds"  
/db\_xref="taxon:9823"  
/sex="male and female"  
/clone\_lib="SCF - porcine spleen"  
/dev\_stage="adult"  
/note="Organ: spleen"  
<1..>847

BASE COUNT 205 a 249 c 200 g 193 t

STS  
ORIGIN

Query Match 3.3%; Score 47; DB 11; Length 847;  
Best Local Similarity 100.0%; Pred. No. 4.3e-17;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 873 GCAGCACAGTGTGGACGCCACCATGAGAGAGAGCAAGTTATGGCCA 919

Db 532 GCAGCACAGTGTGGACGCCACCATGAGAGAGAGCAAGTTATGGCCA 578

## RESULT 10

AF069988

LOCUS

AF069988 1338 bp mRNA linear ROD 23-JUL-1998

DEFINITION Mus musculus nitrilase 1 (Ntl1) mRNA, complete cds.

ACCESSION AF069988

VERSION AF069988.1 GI:3228667

KEYWORDS

SOURCE Mus musculus.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1338)

AUTHORS Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,

Tillib, S., Draganescu, A., Wernuth, P., Rothman, J.H., Huebner, K.,

Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

Nitrilase and Fhit homologs are encoded as fusion proteins in

Drosophila melanogaster and Caenorhabditis elegans

Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

JOURNAL 98337986

MEDLINE 9671749

PUBMED 2 (bases 1 to 1338)

REFERENCE Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,

Tillib, S., Draganescu, A., Wernuth, P., Rothman, J., Huebner, K.,

Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

Direct Submission

TITLE Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson

Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

FEATURES

source

Location/Qualifiers

1..1338

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="1"

/map="1q21-q23"

1..1338

/gene="Ntl1"

58..1029

```
/gene="Nitrilase 1"
/codon_start=1
/product="nitrilase 1"
/protein_id="AAH21634.1"
/db_xref="GI:18204913"
/db_xref="LocusID:27045"
/translaton="MLGFTIRPPHQLLCTGYRLRTPVLCQPRPTMSSSTSWELPL
VAVCVSTPNKQENFKTCAELVQEAARLGAFLPEAFDFIARNPAETLLSLEPLN
GDLGQVSQLARECGIWLISLGFHERGQDWQONKIYNCHVLLNSKGSVVASVYRKTHL
CDVEIPGOGPMRESNYTKPGCTLEPPVKTAGKVLCAICYDMRPELSLKLQAQAEI
LTYPSAGSVYTGPAHWEVLLRARAIESOCYVIAAAQCGRRHETRASVGHWSVVDPMGT
VVAACSEPGGLCLARIDLHFLQOMRQHLVFOHRRPDLVSLGHPLS"
BASE COUNT      347 a   335 c   330 g   326 t
ORIGIN
Query Match      2.3%: Score 33: DB 10: Length 1338;
Best Local Similarity 100.0%: Pred. No. 2.3e-08;
Matches 33: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 1343 TGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAA 1375
|||||
Db 1269 TGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAA 1301
|||||
RESULT 11
BC021634
LOCUS
DEFINITION
Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,
complete cds.
ACCESSION
BC021634
VERSION
BC021634.1 GI:18204912
KEYWORDS
MGC.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1365)
Direct Submission
Submitted (14-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 18 Row: e Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6754855.
Location/Qualifiers
1..1365
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZECB II"
/clone="MGC:13825 IMAGE:4008543"
/tissue_type="Mammary tumor metastatized to lung.
MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
enhancer."
/clone_lib="NCI CGAP_Lu30"
/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6"
59..1030
/codon_start=1
CDS
source
```

```
/product="nitrilase 1"
/protein_id="AAH21634.1"
/db_xref="GI:18204913"
/db_xref="LocusID:27045"
/translaton="MLGFTIRPPHQLLCTGYRLRTPVLCQPRPTMSSSTSWELPL
VAVCVSTPNKQENFKTCAELVQEAARLGAFLPEAFDFIARNPAETLLSLEPLN
GDLGQVSQLARECGIWLISLGFHERGQDWQONKIYNCHVLLNSKGSVVASVYRKTHL
CDVEIPGOGPMRESNYTKPGCTLEPPVKTAGKVLCAICYDMRPELSLKLQAQAEI
LTYPSAGSVYTGPAHWEVLLRARAIESOCYVIAAAQCGRRHETRASVGHWSVVDPMGT
VVAACSEPGGLCLARIDLHFLQOMRQHLVFOHRRPDLVSLGHPLS"
BASE COUNT      368 a   334 c   331 g   332 t
ORIGIN
Query Match      2.3%: Score 33: DB 10: Length 1365;
Best Local Similarity 100.0%: Pred. No. 2.3e-08;
Matches 33: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 1343 TGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAA 1375
|||||
Db 1270 TGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAA 1302
|||||
RESULT 12
AF069985
LOCUS
DEFINITION
Mus musculus nitrilase homolog 1 (Nitr1) gene, alternatively spliced
product, complete cds.
ACCESSION
AF069985
VERSION
AF069985.1 GI:3242979
KEYWORDS
MGC.
SOURCE
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4481)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Werbuth,P., Rothman,J.H., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrilase and Flit homologs are encoded as fusion proteins in
Drosophila melanogaster and Caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
98337986
PUBMED
9671749
2 (bases 1 to 4481)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,
Tillib,S., Draganescu,A., Werbuth,P., Rothman,J., Huebner,K.,
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
Location/Qualifiers
1..4481
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/map="1q21-q23"
541..3944
/gene="Nitr1"
/join(541..606,1177..1263,1638..1889,2015..2118,2362..2495,
2626..2751,3392..3944)
/organism="Nitr1"
/product="nitrilase homolog 1"
/notes="alternatively spliced"
541..606
/gene="Nitr1"
/notes="alternatively spliced"
2626..2751,3392..3658)
/join(605..606,1177..1263,1638..1889,2015..2118,2362..2495,
2626..2751,3392..3658)
/gene="Nitr1"
/notes="alternatively spliced"
59..1030
/codon_start=1
/product="nitrilase homolog 1"
CDS
source
```



\* 6815 6914: gap of unknown length  
\* 6915 8099: contig of 1185 bp in length  
\* 8100 8199: gap of unknown length  
\* 8200 9786: contig of 1587 bp in length  
\* 9787 9886: gap of unknown length  
\* 9887 11574: contig of 1688 bp in length  
\* 11575 12816: gap of unknown length  
\* 11675 12816: contig of 1142 bp in length  
\* 12817 12916: gap of unknown length  
\* 12917 14067: contig of 1151 bp in length  
\* 14068 14167: gap of unknown length  
\* 14168 15370: contig of 1203 bp in length  
\* 15371 15470: gap of unknown length  
\* 15471 16926: contig of 1456 bp in length  
\* 16927 17026: gap of unknown length  
\* 17027 18088: contig of 1062 bp in length  
\* 18089 18188: gap of unknown length  
\* 18189 19315: contig of 1127 bp in length  
\* 19316 19415: gap of unknown length  
\* 19416 20772: contig of 1357 bp in length  
\* 20773 20872: gap of unknown length  
\* 20873 21937: contig of 1065 bp in length  
\* 21938 22037: gap of unknown length  
\* 22038 23147: contig of 1110 bp in length  
\* 23148 23247: gap of unknown length  
\* 23248 24803: contig of 1556 bp in length  
\* 24804 24903: gap of unknown length  
\* 24904 26111: contig of 1208 bp in length  
\* 26112 26211: gap of unknown length  
\* 26212 27647: contig of 1436 bp in length  
\* 27648 27747: gap of unknown length  
\* 27748 30257: contig of 2510 bp in length  
\* 30258 30357: gap of unknown length  
\* 30358 31725: contig of 1368 bp in length  
\* 31726 31825: gap of unknown length  
\* 31826 32912: contig of 1087 bp in length  
\* 32913 33012: gap of unknown length  
\* 33013 34593: contig of 1581 bp in length  
\* 34594 34693: gap of unknown length  
\* 34694 36338: contig of 1645 bp in length  
\* 36339 36438: gap of unknown length  
\* 36439 38139: contig of 1701 bp in length  
\* 38140 38239: gap of unknown length  
\* 38240 39349: contig of 1110 bp in length  
\* 39350 39449: gap of unknown length  
\* 39450 40968: contig of 1519 bp in length  
\* 40969 41068: gap of unknown length  
\* 41069 42068: contig of 1000 bp in length  
\* 42069 42168: gap of unknown length  
\* 42169 43869: contig of 1701 bp in length  
\* 43870 46141: contig of 2172 bp in length  
\* 46142 46241: gap of unknown length  
\* 46242 48158: contig of 1917 bp in length  
\* 48159 48258: gap of unknown length  
\* 48259 50613: contig of 2355 bp in length  
\* 50614 50713: gap of unknown length  
\* 50714 52725: contig of 2012 bp in length  
\* 52726 52825: gap of unknown length  
\* 52826 54534: contig of 1709 bp in length  
\* 54535 54634: gap of unknown length  
\* 54635 56455: contig of 1821 bp in length  
\* 56456 56555: gap of unknown length  
\* 56556 58494: contig of 1939 bp in length  
\* 58495 58594: gap of unknown length  
\* 58595 60496: contig of 1902 bp in length  
\* 60497 60596: gap of unknown length  
\* 60597 62628: contig of 2032 bp in length  
\* 62629 62728: gap of unknown length  
\* 62729 65060: contig of 2332 bp in length  
\* 65061 65160: gap of unknown length  
\* 65161 66543: contig of 1383 bp in length  
\* 66544 66643: gap of unknown length

\* 66644 68412: contig of 1769 bp in length  
\* 68413 68512: gap of unknown length  
\* 68513 71585: contig of 3073 bp in length  
\* 71586 71685: gap of unknown length  
\* 71686 74575: contig of 2890 bp in length  
\* 74576 74675: gap of unknown length  
\* 74676 77865: contig of 3190 bp in length  
\* 77866 77965: gap of unknown length  
\* 77966 80442: contig of 2477 bp in length  
\* 80443 80542: gap of unknown length  
\* 80543 83108: contig of 2566 bp in length  
\* 83109 83208: gap of unknown length  
\* 83209 86895: contig of 3687 bp in length  
\* 86896 86995: gap of unknown length  
\* 86996 91038: contig of 4043 bp in length  
\* 91039 91138: gap of unknown length  
\* 91139 94412: contig of 3274 bp in length  
\* 94413 94512: gap of unknown length  
\* 94513 96959: contig of 2447 bp in length  
\* 96960 97059: gap of unknown length  
\* 97060 98988: contig of 1929 bp in length  
\* 98989 99088: gap of unknown length  
\* 99089 101633: contig of 2545 bp in length  
\* 101634 101733: gap of unknown length  
\* 101734 105899: contig of 4166 bp in length

Query Match 2.3%; Score 33; DB 2; Length 181583;

Best Local Similarity 100.0%; Pred. No. 5.2e-08; Mismatches 0; Indels 0; Gaps 0;

Oy 1343 TGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAA 1375  
|||||

Db 135691 TGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAA 135723  
|||||

#### RESULT 14

AC084821 LOCUS 211772 bp DNA linear ROD 19-FEB-2002  
DEFINITION Mus musculus chromosome 1 clone rp23-395h6, complete sequence.  
ACCESSION AC084821  
VERSION AC084821.25 GI:18702388  
KEYWORDS HTG.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 211772)  
AUTHORS Deschamps,S., Gu,W. and Roe,B.A.  
TITLE Mus musculus BAC Clone rp23-395h6  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 211772)  
AUTHORS Deschamps,S., Gu,W. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL  
REFERENCE 3 (bases 1 to 211772)  
AUTHORS Deschamps,S., Gu,W. and Roe,B.A.  
TITLE Submitted (19-FEB-2002) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
JOURNAL  
COMMENT On Feb 19, 2002 this sequence version replaced gi:18390260.  
Center: Department of Chemistry And Biochemistry  
The University of Oklahoma  
Center code:UOKNOR  
FEATURES  
source Location/Qualifiers  
1..211772  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/Chromosome="1"

BASE COUNT 55607 a 48743 c 49203 g 58215 t 4 others  
ORIGIN /clone="rp23-395hc"  
/clone\_lib="RPCI Mouse BAC Library 23"  
55607 a 48743 c 49203 g 58215 t 4 others

Query Match 2.3%; Score 33; DB 10; Length 211772;  
Best Local Similarity 100.0%; Pred. No. 5.3e-08;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TGCTGAGGCTCAGCAGCACTGGCATTGAAGAA 1375  
|||||  
Db 3237 TGCTGAGGCTCAGCAGCACTGGCATTGAAGAA 3269

BASE COUNT 55806 a 49340 c 51162 g 58732 t 3 others  
ORIGIN /clone\_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"  
55806 a 49340 c 51162 g 58732 t 3 others

Query Match 2.3%; Score 33; DB 10; Length 215043;  
Best Local Similarity 100.0%; Pred. No. 5.3e-08;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TGCTGAGGCTCAGCAGCACTGGCATTGAAGAA 1375  
|||||  
Db 87663 TGCTGAGGCTCAGCAGCACTGGCATTGAAGAA 87631

RESULT 15  
AC087229/c  
LOCUS AC087229 215043 bp DNA linear ROD 14-AUG-2002  
DEFINITION Mus musculus chromosome 1 clone rp23-191a19 strain C57BL/6J,  
complete sequence.  
ACCESSION AC087229  
VERSION AC087229.24 GI:22094398  
KEYWORDS HTG  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 215043)  
Jiang,X., Song,L., Gu,W. and Roe,B.A.  
Mus musculus Chromosome 1 BAC Clone rp23-191a19  
Unpublished  
2 (bases 1 to 215043)  
Jiang,X., Song,L., Gu,W. and Roe,B.A.  
Direct Submission  
Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA  
3 (bases 1 to 215043)  
Jiang,X., Song,L., Gu,W. and Roe,B.A.  
Direct Submission  
Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA  
4 (bases 1 to 215043)  
Jiang,X., Song,L., Gu,W. and Roe,B.A.  
Direct Submission  
Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA  
5 (bases 1 to 215043)  
Jiang,X., Song,L., Gu,W. and Roe,B.A.  
Direct Submission  
Submitted (13-AUG-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA  
6 (bases 1 to 215043)  
Jiang,X., Song,L., Gu,W. and Roe,B.A.  
Direct Submission  
Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA  
On Aug 3, 2002 this sequence version replaced gi:22038565.  
-----  
Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----

FEATURES  
source  
Location/Qualifiers  
1..215043  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/clone="rp23-191a19"

Search completed: December 14, 2002, 18:13:49  
Job time : 4176 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 13:59:05 ; Search time 263 Seconds  
(without alignments)  
12124.827 Million cell updates/sec

Title: US-09-357-675C-1  
Perfect score: 1416  
Sequence: 1 gccactcgtcgcgctnt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :				N_Geneseq_101002:*	
1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*				
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*				
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*				
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*				
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*				
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*				
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*				
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*				
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*				
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*				
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*				
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*				
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*				
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*				
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*				
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*				
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*				
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*				
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*				
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*				
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*				
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*				
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*				
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1314	92.8	1416	21 AAZ46101 Nucleotide sequenc
2	1160	81.9	1382	21 AAF18257 Human prostate can
3	766	54.1	1203	20 AAX30398 DNA encoding a hum
c 4	326	23.0	377	24 ABK39058 cDNA encoding lung
5	91	6.4	505	22 AAH07371 Human cDNA clone (
6	87	6.1	592	22 AAH11324 Human cDNA clone (
c 7	87	6.1	2261	22 ARL14367 Human cDNA sequenc
c 8	87	6.1	2261	22 ABK46133 cDNA encoding colo
9	58	4.1	60	24 ABL39628 Human spliced tran

10	52	3.7	539	22	AAK61236	Human immune/haema
11	30	2.1	3300	20	AAZ11915	Human potassium ch
12	29	2.0	80	24	ABA82722	Human protective D
13	29	2.0	223	23	ABV08728	Human prostate exp
c 14	29	2.0	311	24	ABL82990	Human ovarian canc
c 15	29	2.0	384	23	ABV13242	Human prostate exp
c 16	29	2.0	406	23	ABV04073	Human prostate exp
c 17	29	2.0	416	23	ABV34361	Human prostate exp
18	29	2.0	417	23	ABV58938	Human prostate exp
19	29	2.0	528	23	ABV38626	Human prostate exp
20	29	2.0	563	20	AAV98006	Human secreted pro
c 21	29	2.0	564	24	ABN93758	Gene #256 used to
c 22	29	2.0	748	22	AAK63693	Human immune/haema
23	29	2.0	807	24	ABA82714	Human protective D
24	29	2.0	1193	13	AAQ30670	GGF2BPP2.CDS. Syn
25	29	2.0	1193	15	AAQ58303	GGF-II cDNA sequen
26	29	2.0	1193	15	AAQ62849	GGF-II cDNA sequen
27	29	2.0	1193	16	AAQ74885	Putative bovine gl
28	29	2.0	1193	17	AAT48079	Bovine neurogulin
29	29	2.0	1193	17	AAT30997	Bovine glial growt
30	29	2.0	1193	17	AAT06703	Bovine glial growt
31	29	2.0	1211	22	AAH31352	Human secreted pro
32	29	2.0	1608	21	AAC95501	Human secreted pro
33	29	2.0	2317	22	AAQ03917	NOV9 coding sequen
34	29	2.0	3625	23	AAF87120	Human HLIG-1 cDNA.
35	29	2.0	4843	20	AAV37553	Human ovarian canc
c 36	28	2.0	319	24	ABL79372	Human prostate exp
37	28	2.0	324	23	ABV49621	Human prostate exp
38	28	2.0	326	23	ABV19858	Human prostate exp
39	28	2.0	428	23	ABV17820	Human prostate exp
40	28	2.0	469	23	ABV47612	Human prostate exp
c 41	28	2.0	619	22	AAV22931	Human breast cance
c 42	28	2.0	711	22	AAI97858	Human neuroblastom
43	28	2.0	941	24	ABN98591	Arabidopsis thalia
44	28	2.0	1108	13	AAQ30671	GGF2BPP3.CDS. Syn
45	28	2.0	1108	15	AAQ58304	GGF-II cDNA sequen

ALIGNMENTS

RESULT 1	
AAZ46101	
ID	AAZ46101 standard; DNA; 1416 BP.
XX	AC
AC	AAZ46101;
XX	XX
XX	05-MAY-2000 (first entry)
DT	Nucleotide sequence of the coding region of NIT1 gene.
DE	
XX	NIT1 gene; nitriase; tumour suppressor gene; PHIT; chromosome 3p14.2;
KW	FRA3B; cancer; genome allele inactivation; ss.
XX	
OS	Homo sapiens.
OS	Mus sp.
OS	Drosophila melanogaster.
XX	Caenorhabditis elegans.
XX	
Key	Location/Qualifiers
FT	CDS
FT	3..1415
FT	/*tag= a
FT	/product= (pos: 18..20, aa: Xaa)
FT	/product= (pos: 1179..1181, aa: Gln)
FT	/product= (pos: 1182..1184, aa: Ala)
FT	/note= "contains 7 internal stop codons; Xaa is an unspecified amino acid"
XX	
PN	WO200003685-A2.
XX	
PD	27-JAN-2000.
XX	
PF	20-JUL-1999; 99WO-US16366.

XX 20-JUL-1998; 98US-0093350.  
XX (UYJE-) UNIV JEFFERSON THOMAS.  
XX Croce CM;  
XX  
XX WPI; 2000-1711195/15.  
XX P-PSDB; AAY68739.  
XX  
XX Novel nitrlase homologs used as diagnostic and therapeutic reagents  
XX for the detection and treatment of cancer  
XX  
XX Claim 6; Fig 6; 25pp; English.  
XX  
XX The present sequence represents the coding region of human, murine,  
XX Drosophila melanogaster and Caenorhabditis elegans N1rl gene. The  
XX human and mouse N1rl genes are members of an uncharacterised  
XX mammalian gene family with homology to bacterial and plant nitrilases.  
XX The tumour suppressor gene FHIT in D. melanogaster and C. elegans code  
XX for fusion proteins in which the Phit domain is fused with a Nit domain.  
XX In mouse and humans, FHIT and N1rl are encoded by two different genes,  
XX localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The  
XX human FHIT gene at chromosome 3p14.2, spanning the constitutive  
XX chromosomal fragile site FRA3B, is often altered in most common forms  
XX of human cancer. The N1rl protein overcomes the mutated inactivation  
XX of the genome alleles. The N1rl genes, encoded polypeptides, derivatives  
XX and analogues of them, and antibodies are used as diagnostic and  
XX therapeutic reagents for the detection and treatment of cancers.  
XX  
XX Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;  
XX  
Query Match 92.8%; Score 1314; DB 21; Length 1416;  
Best Local Similarity 99.9%; Pred No. 0;  
Matches 1414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCCACTCGCTCGGCGCTNTCTGGTCCAGACCGGCCCTCCGGATCGGACCCCTCGGAATGG 60  
DB 1 GCCCACTCGCTCGGCGCTNTCTGGTCCAGACCGGCCCTCCGGATCGGACCCCTCGGAATGG 60  
QY 61 TTTTGGCTATATCTTCATGTAGACCTACTCCCTATCCCGTGGCGCGGCTGGGCTTCA 120  
DB 61 TTTTGGCTATATCTTCATGTAGACCTACTCCCTATCCCGTGGCGCGGCTGGGCTTCA 120  
QY 121 TCACAGGCGCTCTCCACAGATTCTGTCCCTCTGTGTCTCGGACTCCGGATACCTCAAC 180  
DB 121 TCACAGGCGCTCTCCACAGATTCTGTCTCCCTCTGTGTCTCGGACTCCGGATACCTCAAC 180  
QY 181 TCTAGTACTTTGTGCTCAGCCAGCCCGCAGAGCCATGGCTATCTCTCTCTCTCTGG 240  
DB 181 TCTAGTACTTTGTGCTCAGCCAGCCCGCAGAGCCATGGCTATCTCTCTCTCTCTGG 240  
QY 241 AACTGCCCCGTGGTGTGTCAGAGTAACTGTCAGGCTAACTGACGCGCAGACAGCAAGAACTTTA 300  
DB 241 AACTGCCCCGTGGTGTGTCAGAGTAACTGTCAGGCTAACTGACGCGCAGACAGCAAGAACTTTA 300  
QY 301 AAACATGTGCTGAGCTGGTTCAGAGGCTGCGAGACTGGTGGCTGCTTCTCTCTCTCTG 360  
DB 301 AAACATGTGCTGAGCTGGTTCAGAGGCTGCGAGACTGGTGGCTGCTTCTCTCTCTCTG 360  
QY 361 CTGAGGCAATTTGACTTCAATTCACGCGGACCCCTGCGAGAGCGCTACACCTGTCTGAACAC 420  
DB 361 CTGAGGCAATTTGACTTCAATTCACGCGGACCCCTGCGAGAGCGCTACACCTGTCTGAACAC 420  
QY 421 TGGGTGGGAACTTTTGGAGAAATACACCCAGCTTGGCCAGGGAATGTGGACTCTGGCTGT 480  
DB 421 TGGGTGGGAACTTTTGGAGAAATACACCCAGCTTGGCCAGGGAATGTGGACTCTGGCTGT 480  
QY 481 CCTTGGGTGGTTCATGAGCTGGCGAAGACTGGGAGGAGCTCAGAAAACTACAATT 540  
DB 481 CCTTGGGTGGTTCATGAGCTGGCGAAGACTGGGAGGAGCTCAGAAAACTACAATT 540  
QY 541 GTCACGTGCTGCTGAACAGCAAGGGGCGAGTAGTGGCCACTTACAGGAAGACACATCTGT 600

DB 541 GTCACGTGCTGCTGAACAGCAAGGGGCGAGTAGTGGCCACTTACAGGAAGACACATCTGT 600  
QY 601 GTGACGTAGAGATTCAGGGCAGGGCCCTATGTGTGAAGCAACTCTACCATGCTGCTGGC 560  
DB 601 GTGACGTAGAGATTCAGGGCAGGGCCCTATGTGTGAAGCAACTCTACCATGCTGCTGGC 560  
QY 661 CCAGCTTTGAGTCACCTGTGACACACAGCAGGCAAGATTGGTCTAGTCTGCTCTGCTATG 720  
DB 661 CCAGCTTTGAGTCACCTGTGACACACAGCAGGCAAGATTGGTCTAGTCTGCTCTGCTATG 720  
QY 721 ACATCGGTTCCCTGAACTCTCTCTGGCATTTGGCTCAAGCTGGAGCAGAGATFACCTTACCT 780  
DB 721 ACATCGGTTCCCTGAACTCTCTCTGGCATTTGGCTCAAGCTGGAGCAGAGATFACCTTACCT 780  
QY 781 ATCCCTTACAGCTTTTCGATCCATTACAGGCCAGCCAGCCACTGGGAGGTGTTGCTGCGGGGCC 840  
DB 781 ATCCCTTACAGCTTTTCGATCCATTACAGGCCAGCCAGCCACTGGGAGGTGTTGCTGCGGGGCC 840  
QY 841 GTGCTATCGAAACCCAGTGTCTATGTAGTGGCAGCAGCAGTGTGGAGCCACCATGAGA 900  
DB 841 GTGCTATCGAAACCCAGTGTCTATGTAGTGGCAGCAGCAGTGTGGAGCCACCATGAGA 900  
QY 901 AGAGAGCAAGTTATGSCCAGCAGCATGTTGGTAGACCCCTGGGGAACAGTGTGGCCCGCT 960  
DB 901 AGAGAGCAAGTTATGSCCAGCAGCATGTTGGTAGACCCCTGGGGAACAGTGTGGCCCGCT 960  
QY 961 GCTCTGAGGGGCCAGGCGCTGCTTGGCCGGAATAGACCTCAACTATCTGCGACAGTTGC 1020  
DB 961 GCTCTGAGGGGCCAGGCGCTGCTTGGCCGGAATAGACCTCAACTATCTGCGACAGTTGC 1020  
QY 1021 GCGGACACTGCTGCTGTTCAGACACCGCAGCCCTGACTCTATGCGCAATCTGGGTCAAC 1080  
DB 1021 GCGGACACTGCTGCTGTTCAGACACCGCAGCCCTGACTCTATGCGCAATCTGGGTCAAC 1080  
QY 1081 CACTCTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCTCCACCCACCCCTGCTGCC 1140  
DB 1081 CACTCTCTTAAGACTTGACTTCTGTGAGTTTAGACCTGCCCTCCACCCACCCCTGCTGCC 1140  
QY 1141 ACTATGAGCTAGTGTCTCATGTGACTTGGAGGAGGATCCAGGCTCCCTCTCACCTG 1200  
DB 1141 ACTATGAGCTAGTGTCTCATGTGACTTGGAGGAGGATCCAGGCTCCCTCTCACCTG 1200  
QY 1201 GAGAACCTTGACTCTCTTGTATGGAACACAGATGGCTGCTGGGAAAGAACTTTCACCT 1260  
DB 1201 GAGAACCTTGACTCTCTTGTATGGAACACAGATGGCTGCTGGGAAAGAACTTTCACCT 1260  
QY 1261 GAGCTTACCTCAGGTGACACTGACATGCTTCAGAAAGGTGGAAATTTATATAGTCAATTGT 1320  
DB 1261 GAGCTTACCTCAGGTGACACTGACATGCTTCAGAAAGGTGGAAATTTATATAGTCAATTGT 1320  
QY 1321 TATTTTCATGGAACTGAAGTTCTGCTGAGGCTGAGGAGGCTGGCATTGAAAAATATAA 1380  
DB 1321 TATTTTCATGGAACTGAAGTTCTGCTGAGGCTGAGGAGGCTGGCATTGAAAAATATAA 1380  
QY 1381 TAATCATAAAGTCAAAAAAATAAAAAAATAAAAAA 1416  
DB 1381 TAATCATAAAGTCAAAAAAATAAAAAAATAAAAAA 1416

RESULT 2  
AAF16257  
ID AAF16257 standard; cDNA; 1382 BP.  
XX AAF16257;  
AC AAF16257;  
XX  
XX 13-MAR-2001 (first entry)  
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.  
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
XX neuroprotective; cytosolic; cardiolipin; immunomodulatory; muscular;  
XX vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW

KW	antibacterial; gene therapy; neural; immune; reproductive; renal; gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW	wound; infectious disease; ss.
XX	Homo sapiens.
OS	WO200055174-A1.
PN	21-SEP-2000.
XX	
PD	08-MAR-2000; 2000WO-US05988.
XX	
PF	12-MAR-1999; 99US-0124270.
XX	(HUMA-) HUMAN GENOME SCI INC.
PA	(ROSE/) ROSEN C A.
XX	Rosen CA, Ruben SW;
PI	WIPI; 2000-587513/55.
XX	P-PSDB; AAB57054.
DR	
XX	
PT	Prostate cancer associated gene sequences, referred to as prostate
PT	cancer antigens, useful for treatment, prevention, and diagnosis of
PT	disorders such as prostate cancer -
XX	
PS	Claim 1; Page 1124; 2338pp; English.
XX	
CC	AALF5566 to AAF16505 encode the human prostate cancer associated
CC	proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC	The prostate cancer antigens can have neuroprotective, cytostatic,
CC	cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC	nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC	and can be used in gene therapy. The prostate cancer antigen
CC	polynucleotides may be used for detection of prostate cancer, chromosome
CC	identification, as chromosome markers, and for numerous other diagnostic
CC	or research purposes. The prostate cancer antigens may be used to treat
CC	disorders such as neural, immune, muscular, reproductive.
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC	diseases, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC	AAB57303 represent sequences used in the exemplification of the present
CC	invention.
XX	
SQ	Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;
Query Match            81.9%; Score 1160; DB 21; Length 1382;	
Best Local Similarity   99.9%; Pred. No. 0;	
Matches 1280; Conservative   0; Mismatches   0; Indels   1; Gaps   1;	
QY	110 GCTGGGCTTCATCACCAGGCCTCTCACAGATTCTGTCCCTTCTGTGTCCTGGACTCGG 169
Db	
	79 GCTGGGCTTCATCACCAGGCCTCTCACAGATTCTGTCCCTTCTGTGTCCTGGACTCGG 138
QY	170 GATACCTCACTCTCACTACTTTGTGCTCAGCCAGGCCAGGCATGGCTATCTCCTC 229
Db	
	139 GATACCTCACTCTCACTACTTTGTGCTCAGCCAGGCCAGGCATGGCTATCTCCTC 198
QY	230 TTCTCTCTGGAACTGCCCTGGTGCTGTGTGCCAGGTAACTCGAGCCAGACAAGA 289
Db	
	199 TTCTCTCTGGAACTGCCCTGGTGCTGTGTGCCAGGTAACTCGAGCCAGACAAGA 258
QY	290 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGCTGCCT 349
Db	
	259 ACAGAACTTTAAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGCTGCCT 318
QY	350 GGCTTTCTCGCTGAGGCAATTTGACTTATTTGCACGGGACCCCTCGAGAGACGCTACAC 409
Db	
	319 GGCTTTCTCGCTGAGGCAATTTGACTTATTTGCACGGGACCCCTCGAGAGACGCTACAC 378
QY	410 GTCTGAACCACTGGGTGGGAAACATTTTGAAGAATAACCCAGCTTGCACGGGAATGTGG 469
Db	
	379 GTCTGAACCACTGGGTGGGAAACATTTTGAAGAATAACCCAGCTTGCACGGGAATGTGG 438

DE DNA encoding a human secreted protein.

XX Secreted protein; cancer; tumour; neurodegenerative disorder;

KW developmental abnormality; foetal deficiency; blood disorder;

KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;

KW renal disease; diabetes; inflammation; allergy; ischemic shock;

KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;

KW prostate disease; asthma; osteoporosis; arthritis; ss.

XX Homo sapiens.

OS W09907891-Al.

PN 18-FEB-1999.

XX 04-AUG-1998; 98WO-US16235.

XX 19-AUG-1997; 97US-0056732.

PR 05-AUG-1997; 97US-0054798.

PR 05-AUG-1997; 97US-0054803.

PR 05-AUG-1997; 97US-0054804.

PR 05-AUG-1997; 97US-0054806.

PR 05-AUG-1997; 97US-0054807.

PR 05-AUG-1997; 97US-0054808.

PR 05-AUG-1997; 97US-0054809.

PR 05-AUG-1997; 97US-0055309.

PR 05-AUG-1997; 97US-0055310.

PR 05-AUG-1997; 97US-0055312.

PR 05-AUG-1997; 97US-0055386.

PR 05-AUG-1997; 97US-0055311.

PR 18-AUG-1997; 97US-0055970.

PR 18-AUG-1997; 97US-0055986.

PR 19-AUG-1997; 97US-0056365.

PR 19-AUG-1997; 97US-0058366.

PR 19-AUG-1997; 97US-0058557.

PR 19-AUG-1997; 97US-0056370.

PR 19-AUG-1997; 97US-0056371.

PR 19-AUG-1997; 97US-0056563.

PR 19-AUG-1997; 97US-0056731.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;

PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;

XX WPI: 1999-167452/14.

DR P-PSDB; AAT10877.

XX New isolated human genes encoding secreted polypeptides - useful for

PT diagnosis and treatment of pathological diseases

PT Claim 3; Page 265-266; 331pp; English.

XX The specification describes secreted proteins and their corresponding

CC polynucleotides which are useful for preventing, treating or ameliorating

CC medical conditions, e.g. by protein or gene therapy. Pathological

CC conditions can also be diagnosed by determining the amount of the

CC secreted polypeptides in a sample or by determining the presence of

CC mutations in the polynucleotides. Specific uses are described for each

CC of the products, based on which tissues they are most highly

CC expressed in, and include developing products for the diagnosis or

CC treatment of cancer, tumours, neurodegenerative disorders, developmental

CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,

CC diseases of the immune system, autoimmune diseases, hepatic and renal,

CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's

CC and cognitive disorders, schizophrenia, cardiovascular disorders,

CC prostate diseases, asthma, disorders involving osteoclasts such as

CC osteoporosis, arthritis or malignancies, diseases of testes, lung or

CC thymus, digestive/endocrine disorders, infections and AIDS. The

CC polypeptides are also useful for identifying their binding partners.

XX Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

Query Match 54.1%; Score 766; DB 20; Length 1203;

Best Local Similarity 100.0%; Pred. No. 8.4e-305;

Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 CTATGTTGTAAGCAACTTACCATGCGTGGGCCAGTCTTGAGTCACCTGTCAGCACAC 687

DB 161 CTATGTTGTAAGCAACTTACCATGCGTGGGCCAGTCTTGAGTCACCTGTCAGCACAC 220

QY 688 CAGCAGGCAAGATTGGTCTAGCTGTCTATGACATGCGGTTCCTGAACTCTCTCTGG 747

DB 221 CAGCAGGCAAGATTGGTCTAGCTGTCTATGACATGCGGTTCCTGAACTCTCTCTGG 280

QY 748 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCTCTTTCAGCTTTGGATCCATTACAG 807

DB 281 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCTCTTTCAGCTTTGGATCCATTACAG 340

QY 808 GCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTCTATCGAAACCCAGTGTATGTAG 867

DB 341 GCCAGCCCACTGGGAGGTGTTGCTGCGGGCCCGTCTATCGAAACCCAGTGTATGTAG 400

QY 868 TGGCAGCAGCAGATGTGGACGCCACCATGAGAAGAGACAAGTTATGCCACAGCATGG 927

DB 401 TGGCAGCAGCAGATGTGGACGCCACCATGAGAAGAGACAAGTTATGCCACAGCATGG 460

QY 928 TGGTAGACCCCTGGGGAACAGTGGTGGCCCGTCTCTGAGGGGCCAGCCCTCTGCCTTG 987

DB 461 TGGTAGACCCCTGGGGAACAGTGGTGGCCCGTCTCTGAGGGGCCAGCCCTCTGCCTTG 520

QY 988 CCCGAATAGACCTCAACTATCTGCGACAGTTGCGCGGACACCTGCTGTGTTCAGCAC 1047

DB 521 CCCGAATAGACCTCAACTATCTGCGACAGTTGCGCGGACACCTGCTGTGTTCAGCAC 580

QY 1048 GCAGGCCCTGACCTCTATGCAATCTGGGTCACCCACTGTCTTAAGACTTTGACTTCTGTGA 1107

DB 581 GCAGGCCCTGACCTCTATGCAATCTGGGTCACCCACTGTCTTAAGACTTTGACTTCTGTGA 640

QY 1108 GTTTAGACCTGCCCTCCACCCACCCCTGCCACTATGAGCTAGTGTCTGTGACTTG 1167

DB 641 GTTTAGACCTGCCCTCCACCCACCCCTGCCACTATGAGCTAGTGTCTGTGACTTG 700

QY 1168 GAGGAGGATCAGGCACAGCTCCCTCTCACTTGGAGAACCTTGACTCTCTTGATGGAACA 1227

DB 701 GAGGAGGATCAGGCACAGCTCCCTCTCACTTGGAGAACCTTGACTCTCTTGATGGAACA 760

QY 1228 CAGATGGGCTGCTGGGAAGAACTTTTACCTGAGCTTCACCTGAGGTCAGACTGCAGT 1287

DB 761 CAGATGGGCTGCTGGGAAGAACTTTTACCTGAGCTTCACCTGAGGTCAGACTGCAGT 820

QY 1288 TTTCAGAAAGGTGGAATTTTATATAGTTCATGTTTATTTTCATGAAACTGAAGTCTCTGCTG 1347

DB 821 TTTCAGAAAGGTGGAATTTTATATAGTTCATGTTTATTTTCATGAAACTGAAGTCTCTGCTG 880

QY 1348 AGGGCTGAGCAGCACTGGCAATTTGAAATAATATATATATATAAGTC 1393

DB 881 AGGGCTGAGCAGCACTGGCAATTTGAAATAATATATATATAAGTC 936

RESULT 4

ABK39058/c

ID ABK39058 standard; cDNA; 377 BP.

XX

AC ABK39058;

XX

DT 21-MAY-2002 (first entry)

XX

DE CDNA encoding lung tumour protein clone R0130:B11.

XX

KW Lung tumour; cancer; T cell; immune response stimulator;

KW cytostatic; gene; ss.

XX

OS Homo sapiens.

XX

PN W0200204514-A2.

```
XX PD 17-JAN-2002.
XX PF 10-JUL-2001; 2001WO-US22058.
XX PR 11-JUL-2000; 2000US-0614124.
XX PR 29-AUG-2000; 2000US-0631563.
XX PR 08-SEP-2000; 2000US-0658824.
XX PR 26-SEP-2000; 2000US-0671325.
XX PR 06-OCT-2000; 2000US-0677419.
XX PR 30-OCT-2000; 2000US-0702705.
XX PR 13-DEC-2000; 2000US-0736457.
XX PR 03-MAY-2001; 2001US-0849626.
XX PA (CORI-) CORIXA CORP.
XX PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
XX PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
XX PI McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;
XX WPI; 2002-164634/21.
XX DR
XX PT Novel polynucleotide encoding a lung tumour polypeptide useful for
XX PT stimulating and/or expanding T cells specific for a tumour protein
XX PS Claim 1; SEQ ID NO 1096; 223pp; English.
XX CC The invention describes an isolated polynucleotide and polypeptide
XX CC useful for stimulating and/or expanding T cells specific for a tumour
XX CC protein for determining the presence of a cancer in a patient. A
XX CC composition containing the polynucleotide and/or polypeptide is useful
XX CC for treating a lung cancer in a patient. The polypeptide is useful for
XX CC removing tumour cells from a biological sample. The polynucleotide is
XX CC also useful as probe or primer to detect the level of mRNA encoding a
XX CC tumour protein. This sequence encodes a lung tumour associated protein
XX CC or protein fragment, described in the method of the invention.
XX CC Note: the sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 377 BP; 94 A; 100 C; 112 G; 71 T; 0 other;

Query Match 23.0%; Score 326; DB 24; Length 377;
Best Local Similarity 99.7%; Pred. No. 3.7e-124;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 129 CTTCTCCTCAGAGATTCCTGCTCCCTCTCTGTCTGGACTCGGATACCTCAACTCTCAGTA 188
DB 377 CTTCTCCTCAGAGATTCCTGCTCCCTCTCTGTCTGGACTCGGATACCTCAACTCTCAGTA 318
QY 189 CTTTGTGCTCAGCCAGGCGCCAGAGCATGGCTATCTCTCTCTCTGCGAAGCTGCC 248
DB 317 CTTTGTGCTCAGCCAGGCGCCAGAGCATGGCTATCTCTCTCTCTGCGAAGCTGCC 258
QY 249 CTGGTGGCTGTGTGCCAGGTAAATCATCGACGCCAGACAAGCAAGCAACATTTAAACATGT 308
DB 257 CTGGTGGCTGTGTGCCAGGTAAATCATCGACGCCAGACAAGCAACATTTAAACATGT 198
QY 309 GCTGAGCTGGTTTCAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 368
DB 197 GCTGAGCTGGTTTCAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 138
QY 369 TTTGACTTTCATTTCAGGGGACCTCGCAGAGACGCTACACCTGCTGCTGAACACACTGGGTGG 428
DB 137 TTTGACTTTCATTTCAGGGGACCTCGCAGAGACGCTACACCTGCTGCTGAACACACTGGGTGG 78
QY 429 AAACCTTTTGAAGAATACACCCAGCTTGGCAGGGAATGTGGACTGTGGCTGTCTTGGGT 488
DB 77 AAACCTTTTGAAGAATACACCCAGCTTGGCAGGGAATGTGGACTGTGGCTGTCTTGGGT 18
QY 489 GGTTCCTCATGAGCGTGG 505
|||||
```

```
DB 17 GGTTCCTCATGAGCGTGG 1
RESULT 5
AAH07371
ID AAH07371 standard; cDNA; 505 BP.
XX AC AAH07371;
XX DT 26-JUN-2001 (first entry)
XX XX Human cDNA clone (5'-primer) SEQ ID NO:4206.
XX DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX KW Homo sapiens.
XX OS
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs.
XX PS Claim 1; SEQ ID 4206; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesizing 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 505 BP; 105 A; 120 C; 167 G; 108 T; 5 other;

Query Match 6.4%; Score 91; DB 22; Length 505;
Best Local Similarity 100.0%; Pred. No. 1e-27;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TCTGGCTCCAGACGCCCTCCGGATCGGACCTCGGAATGGTTTGGCTATATCTTCATG 79
```

```
|||||
Db 19 TCTGGCTCAGACCGCCCTCCGGATCGGACCGCGAATGGTTTGGCTATATCTTCATG 78
QY 80 TAGGACCTACTCCCTATCCCGTCCGCGCGGG 110
Db 79 TAGGACCTACTCCCTATCCCGTCCGCGCGGG 109

RESULT 6
AAH11324
ID AAH11324 standard; cDNA; 592 BP.
AC AAH11324;
XX
XX
XX 26-JUN-2001 (first entry)
XX Human cDNA clone (3'-primer) SEQ ID NO:8159.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 95JP-0248036.
XX 27-AUG-1999; 95JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
XX
XX Claim 3: SEQ ID 8159; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX of the present invention.
XX
XX Sequence 592 BP; 164 A; 134 C; 138 G; 148 T; 8 other;
```

```
Query Match 6.1%; Score 87; DB 22; Length 592;
Best Local Similarity 100.0%; Pred. No. 4.5e-26;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 ATATAGTCATTTGTTTATTTTCATGGAACCTGAAGTTCTGTGAGGGCTGAGCAGCACTGGC 1366
|||||
Db 1 ATATAGTCATTTGTTTATTTTCATGGAACCTGAAGTTCTGTGAGGGCTGAGCAGCACTGGC 60

QY 1367 ATTGAAAAATATAATAATCATAAAGTC 1393
|||||
Db 61 ATTGAAAAATATAATAATCATAAAGTC 87

RESULT 7
AAH14367/c
ID AAH14367 standard; cDNA; 2261 BP.
AC AAH14367;
XX
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:11772.
DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 95JP-0248036.
XX 27-AUG-1999; 95JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
XX
XX Claim 8: SEQ ID 11772; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX of the present invention.
```

CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX SQ Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;

Query Match 6.1%; Score 87; DB 22; Length 2261;

Best Local Similarity 100.0%; Pred. No. 3.9e-26;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 ATATAGTCATTCTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGCGTGAGCAGCACTGGC 1366

Db 2261 ATATAGTCATTCTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGCGTGAGCAGCACTGGC 2202

QY 1367 ATTGAAAAATATAATAATCAATAAGTC 1393

Db 2201 ATTGAAAAATATAATAATCAATAAGTC 2175

#### RESULT 8

ABK46133/c

ID ABK46133 standard; cDNA; 2261 BP.

XX AC ABK46133;

XX DT 05-JUN-2002 (first entry)

XX DE cDNA encoding colon tumour protein, SEQ ID No 1684.

XX KW Human; colon tumour; vaccine; colon cancer; immunogenic;

XX KW immunotherapy; gene; ss.

XX OS Homo sapiens.

XX PN WO200212328-A2.

XX PD 14-FEB-2002.

XX PF 31-JUL-2001; 2001WO-US24218.

XX PR 03-AUG-2000; 2000US-223283P.

XX PR 28-MAR-2001; 2001US-279763P.

XX PR 29-JUN-2001; 2001US-302051P.

XX PA (CORI-) CORIXA CORP.

XX PI King GE, Meagher MJ, Xu J, Secrist H;

XX DR WPI; 2002-241739/29.

XX PT New colon cancer polypeptides and polynucleotides, useful as vaccines,  
XX for diagnosing, preventing, and treating colon cancer, and as markers  
XX for the progression of cancer -

XX PS Claim 1; SEQ ID No 1684; 147pp; English.

XX CC The invention relates to polynucleotides encoding colon tumour proteins.  
XX CC The polynucleotides and encoded polypeptides are useful in pharmaceutical  
XX compositions, such as vaccines, for the diagnosis, prevention, and  
XX treatment of colon cancer. Polynucleotide sequences may be used as  
XX hybridisation probes or primers, and in the design and preparation of  
XX ribozyme molecules for inhibiting expression of tumour polypeptides and  
XX proteins in tumour cells. The compositions are useful for stimulating an  
XX immune response against cancer, particularly for the immunotherapy of  
XX colon cancer, and as markers for the progression of cancer.  
XX CC ABK44450-ABK46237 represent coding sequences of human colon tumour  
XX proteins of the invention.  
XX CC Note: With the exception of SEQ ID No 1 and 2, the sequence data  
XX for this patent did not form part of the printed specification but was  
XX supplied by the European Patent Office.

XX SQ Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;

XX Query Match 6.1%; Score 87; DB 24; Length 2261;

Best Local Similarity 100.0%; Pred. No. 3.9e-26;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 ATATAGTCATTCTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGCGTGAGCAGCACTGGC 1366

Db 2261 ATATAGTCATTCTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGCGTGAGCAGCACTGGC 2202

QY 1367 ATTGAAAAATATAATAATCAATAAGTC 1393

Db 2201 ATTGAAAAATATAATAATCAATAAGTC 2175

#### RESULT 9

ABN39628

ID ABN39628 standard; DNA; 60 BP.

XX AC ABN39628;

XX DT 15-JUL-2002 (first entry)

XX DE Human spliced transcript detection oligonucleotide SEQ ID NO:12376.

XX KW Human; mouse; rat; splice transcript; detection; RNA transcript;

XX KW splice variant; transcriptome; oligonucleotide library; ss.

XX OS Homo sapiens.

XX PN WO200210449-A2.

XX PD 07-FEB-2002.

XX PF 20-JUL-2001; 2001WO-IB01903.

XX PR 28-JUL-2000; 2000US-221607P.

XX PR 02-MAY-2001; 2001US-287724P.

XX PA (COMP-) COMPUGEN INC.

XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX DR WPI; 2002-257383/30.

XX PT New oligonucleotide libraries comprising oligonucleotides which  
XX selectively hybridize to mRNAs transcribed from a transcription unit of  
XX a genome, useful for detecting tissue-, pathology-, and  
XX developmental-specific genes -

XX PS Example 1; SEQ ID 12376; 47pp; English.

XX CC The present invention describes oligonucleotide libraries for detecting  
XX messenger RNAs that populate a (sub-)transcriptome, where the  
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
XX transcription units that populate a genome. The library comprises  
XX several oligonucleotides, each capable of hybridising selectively to a  
XX set of messenger RNAs transcribed from a given transcription unit of  
XX the genome, which encodes one or more messenger RNA splice variants.  
XX The oligonucleotide libraries are useful for detecting mRNAs from a  
XX biological sample, in expression profiling studies, in qualitatively or  
XX quantitatively characterising the corresponding transcriptome, and in  
XX detecting RNA transcripts and splice variants of human or animal  
XX transcriptomes. The libraries may also be used as specialised mini  
XX libraries to detect transcripts of a sub-transcriptome under a  
XX particular biological or pathological state, and so allowing the  
XX detection of tissue- and pathology-specific genes such as those genes  
XX only expressed in specific tissue under a specific pathological  
XX condition; to detect developmental specific genes; and to detect RNA  
XX transcripts and splice variants of a transcriptome of a patient suffering  
XX from a particular disorder. ABN27253 to ABN59589 represent  
XX oligonucleotide sequences from rats, humans and mice, which are used in  
XX the exemplification of the present invention.  
XX N.B. The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

```
XX
SQ   Sequence 60 BP; 9 A; 20 C; 14 G; 17 T; 0 Other;

Query Match      4.1%; Score 58; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 4.5e-14;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY   20  TCTGGCTCCAGACCGCCCTCCGACCTGGACCTGGCAATGGTTTGGCTATATCTTCA 77
      |||||||
Db    3  TCTGGCTCCAGACCGCCCTCCGACCTGGACCTGGCAATGGTTTGGCTATATCTTCA 60

RESULT 10
AAK61236
ID   AAK61236 standard; cdna; 539 BP.
XX
AC   AAK61236;
XX
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cdna SEQ ID NO:6295.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
XX
PD   09-AUG-2001.
XX
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
```



```
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-483426/52.
XX P-PSDB; AAM88455.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Claim 1: SEQ ID NO 6296; 307lpp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;
XX
XX Query Match 3.7%; Score 52; DB 22; Length 539;
XX Best Local Similarity 100.0%; Pred. No. 1e-11;
XX Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 109 GGCTGGCTTCATCACAGGCTCCCTCACAGATTCTGTCCTTCGTGTGCC 160
XX
XX 371 GGCTGGCTTCATCACAGGCTCCCTCACAGATTCTGTCCTTCGTGTGCC 422
XX
XX
XX RESULT 11
XX ID AAZ11915
XX ID AAZ11915 standard; cDNA; 3300 BP.
XX AC AAZ11915;
XX
XX 30-NOV-1999 (first entry)
```

```
XX Human potassium channel K+Hnov59 cDNA.
XX
XX Potassium channel; ataxia; arrhythmia; epilepsy; Bartter's syndrome;
XX cardiovascular disorder; CNS disorder; renal disorder; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 50..1285
XX /*tag= a
XX /product= "Human K+Hnov59 potassium channel"
XX
XX WO9943696-A1.
XX
XX 02-SEP-1999.
XX
XX 22-FEB-1999; 99WO-US03826.
XX
XX 19-JAN-1999; 99US-0116448.
XX 25-FEB-1998; 98US-0076687.
XX 07-AUG-1998; 98US-0095836.
XX
XX (AXYS-) AXYS PHARM INC.
XX
XX Curran ME, Hu P, Miller AP, Rutter M, Wang J;
XX
XX WPI: 1999-527591/44.
XX P-PSDB; AAY34133.
XX
XX New nucleic acids encoding mammalian K+Hnov potassium channel
XX proteins, useful for the diagnosis and treatment of episodic ataxia
XX with myokymia, cardiac arrhythmia, epilepsy and Bartter's syndrome
XX
XX Claim 4; Page 102-104; 112pp; English.
XX
XX This sequence represents human potassium channel K+Hnov59 cDNA.
XX K+Hnov proteins have a high degree of homology to known potassium
XX channels and may be alpha subunits, which form the functional channel,
XX or accessory subunits that act to modulate the channel activity. K+Hnov59
XX is a 4 transmembrane domain, 2 pore domain potassium channel. The gene
XX localisation using primers AAZ11939 and AAZ11940. K+Hnov cDNAs
XX were isolated by extension of expressed sequence tags (ESTs) which were
XX related but not identical to known human potassium channels. Potential
XX polymorphisms detected as sequence variants between multiple
XX independent clones. Potassium channels have critical roles in various
XX cell types and biochemical pathways. Defective potassium channels are
XX known to cause four human diseases: episodic ataxia with myokymia;
XX cardiac arrhythmia (long QT syndrome); epilepsy; and Bartter's syndrome.
XX As potassium channels are critical components of virtually all cells,
XX it is likely that abnormal potassium channels are also implicated in
XX certain renal, cardiovascular and central nervous system (CNS)
XX disorders. Nucleotides encoding K+Hnov proteins may be used for
XX identifying homologous or related proteins and the DNA sequences encoding
XX them. They may be used to produce compositions that modulate the
XX expression and function of the K+Hnov protein and in studying the
XX biochemical pathways associated with it. They may also be used for the
XX recombinant production of K+Hnov protein in fermentation cultures.
XX Additionally, such nucleotides may be used in gene therapy protocols for
XX the treatment of diseases associated with abnormal potassium channels.
XX
XX Sequence 3300 BP; 997 A; 629 C; 680 G; 994 T; 0 other;
XX
XX Query Match 2.1%; Score 30; DB 20; Length 3300;
XX Best Local Similarity 100.0%; Pred. No. 0.0093;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1387 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
XX
XX 3238 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 3267
XX
```

```
RESULT 12
ABA82722
ID ABA82722 standard; DNA; 80 BP.
XX
AC ABA82722;
XX
DT 07-FEB-2002 (first entry)
XX
DE Human protective DNA sequence CNI-00735 fragment #8.
XX
KW Human; protective sequence; cell death; cancer; autoimmune disease;
KW neurological disorder; stroke; cytostatic; neuroprotective; gene therapy;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200176457-A2.
XX
PD 18-OCT-2001.
XX
PF 09-APR-2001; 2001WO-US11663.
XX
PR 11-APR-2000; 2000US-0547735.
XX
PA (COGE-) COCENT NEUROSCIENCE INC.
XX
PI Thomas MB, Portbury SD, Puranam K, Katz LC, Lo DC, Barney S;
XX
DR WPI; 2002-025874/03.
XX
DR P-PSDB; ABB44640.
XX
PT New protective sequences and their products, useful for diagnosing and
PT treating diseases involving cell death, including neurological
PT disorders e.g. stroke and for identifying modulators of expression of
PT the protective sequences
XX
PS Claim 2; Fig 5; 283pp; English.
XX
CC The present invention relates to protective sequence proteins
CC (ABA4624-ABBA4830) and their coding sequences (ABA82701-ABA82937).
CC The sequences, when introduced into a cell either predisposed to undergo
CC cell death or in the process of undergoing cell death, prevent, delay or
CC rescue the cell from death, hence, these sequences are named "protective
CC sequences". The sequences are useful for treating and/or ameliorating
CC cancer, autoimmune diseases and neurological disorders e.g. stroke.
CC Further examples of diseases which may be treated by the present
CC invention are given in the specification.
XX
SQ Sequence 80 BP; 40 A; 10 C; 15 G; 15 T; 0 other;
Query Match 2.0%; Score 29; DB 24; Length 80;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 39 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 67
|||||
RESULT 13
ABV08728
ID ABV08728 standard; cDNA; 223 BP.
XX
AC ABV08728;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 8719.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
```

```
XX WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX
PS Claim 1; Page 1376; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 223 BP; 76 A; 44 C; 64 G; 38 T; 1 other;
Query Match 2.0%; Score 29; DB 23; Length 223;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 188 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 216
|||||
RESULT 14
ABL82990/c
ID ABL82990 standard; cDNA; 311 BP.
XX
AC ABL82990;
XX
DT 17-MAY-2002 (first entry)
XX
DE Human ovarian cancer related cDNA clone SEQ ID NO:5968.
XX
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192581-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US17756.
XX
```

```
PR 26-MAY-2000; 2000US-207484P.
XX (CORI-) CORIXA CORP.
XX Algate PA, Harlocker SL, Jones R;
XX WPI; 2002-122075/16.
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
XX polypeptide
XX
XX Claim 1; SEQ ID 5968; 489pp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
XX techniques.
XX
XX Sequence 311 BP; 96 A; 42 C; 45 G; 128 T; 0 other;
SQ
Query Match 2.0%; Score 29; DB 24; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 40 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 12
RESULT 15
ABV13242/c
ID ABV13242 standard; cDNA; 384 BP.
XX
XX AC ABV13242;
XX
XX DT 13-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 13233.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200160860-A2.
XX
XX PD 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
```

```
PR 18-JUL-2000; 2000US-219007P.
XX 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 2192; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 384 BP; 136 A; 69 C; 73 G; 106 T; 0 other;
SQ
Query Match 2.0%; Score 29; DB 23; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 42 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 14
Search completed: December 14, 2002, 17:02:56
Job time : 290 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 14:01:05 ; Search time 59 Seconds  
(without alignments)  
7360.242 Million cell updates/sec

Title: US-09-357-675C-1  
Perfect score: 1416  
Sequence: 1 gccactcgtcgcgctctt.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	2.1	3300	4 US-09-336-643A-82	Sequence 82, Appl
2	29	2.0	1193	1 US-08-036-555B-134	Sequence 134, App
3	29	2.0	1193	1 US-08-469-569-134	Sequence 134, App
4	29	2.0	1193	1 US-08-249-322A-134	Sequence 134, App
5	29	2.0	1193	1 US-08-469-526A-134	Sequence 134, App
6	29	2.0	1193	2 US-08-734-591A-134	Sequence 134, App
7	29	2.0	1193	2 US-08-469-660-134	Sequence 134, App
8	29	2.0	1193	3 US-08-341-018-3	Sequence 3, Appli
9	29	2.0	1193	3 US-08-470-335-134	Sequence 134, App
10	29	2.0	1193	4 US-08-735-021-134	Sequence 134, App
11	29	2.0	1193	4 US-08-734-664A-134	Sequence 134, App
12	29	2.0	1193	4 US-08-470-339-134	Sequence 134, App
13	29	2.0	1193	4 US-08-467-602-134	Sequence 134, App
14	29	2.0	1193	5 PCT-US94-05083C-130	Sequence 130, App
15	29	2.0	1193	5 PCT-US95-06846A-134	Sequence 134, App
16	29	2.0	4843	3 US-08-986-485-1	Sequence 1, Appli
17	28	2.0	1108	1 US-08-036-555B-135	Sequence 135, App
18	28	2.0	1108	1 US-08-469-569-135	Sequence 135, App
19	28	2.0	1108	1 US-08-249-322A-135	Sequence 135, App
20	28	2.0	1108	1 US-08-469-526A-135	Sequence 135, App
21	28	2.0	1108	2 US-08-734-591A-135	Sequence 135, App
22	28	2.0	1108	2 US-08-469-660-135	Sequence 135, App
23	28	2.0	1108	3 US-08-341-018-5	Sequence 5, Appli
24	28	2.0	1108	3 US-08-470-335-135	Sequence 135, App
25	28	2.0	1108	4 US-08-735-021-135	Sequence 135, App
26	28	2.0	1108	4 US-08-734-664A-135	Sequence 135, App
27	28	2.0	1108	4 US-08-470-339-135	Sequence 135, App

ALIGNMENTS

RESULT 1

US-09-336-643A-82  
; Sequence 82, Application US/09336643A  
; Patent No. 6399761  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Andrew P.  
; APPLICANT: Curran, Mark Edward  
; APPLICANT: Hu, Ping  
; APPLICANT: Rutter, Marc  
; APPLICANT: Wang, Jian-Wang  
; TITLE OF INVENTION: No. 6399761el Human Potassium Channels  
; FILE REFERENCE: SEQ-15P  
; CURRENT APPLICATION NUMBER: US/09/336,643A  
; CURRENT FILING DATE: 1999-06-18  
; PRIOR APPLICATION NUMBER: 60/076,687  
; PRIOR FILING DATE: 1998-08-07  
; PRIOR APPLICATION NUMBER: 60/116,448  
; PRIOR FILING DATE: 1999-01-19  
; PRIOR APPLICATION NUMBER: PCT/US99/03826  
; PRIOR FILING DATE: 1999-02-22  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 82  
; LENGTH: 3300  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (50)...(1285)  
US-09-336-643A-82

28	2.0	1108	4	US-08-467-602-135	Sequence 135, App
29	2.0	1108	5	PCT-US94-05083C-131	Sequence 131, App
30	2.0	1108	5	PCT-US95-06846A-135	Sequence 135, App
31	2.0	3238	4	US-08-123-934A-5	Sequence 5, Appli
32	2.0	3238	5	PCT-US94-10080-5	Sequence 5, Appli
33	2.0	8201	1	US-08-253-155A-9	Sequence 9, Appli
34	2.0	989	2	US-08-874-460-1	Sequence 1, Appli
35	2.0	1969	2	US-08-541-033A-7	Sequence 7, Appli
36	2.0	1969	2	US-08-828-451-7	Sequence 7, Appli
37	2.0	2096	2	US-08-541-033A-19	Sequence 19, Appl
38	2.0	2096	2	US-08-828-451-19	Sequence 19, Appl
39	2.0	2099	2	US-08-541-033A-3	Sequence 3, Appli
40	2.0	2099	2	US-08-828-451-3	Sequence 3, Appli
41	2.0	2137	2	US-08-541-033A-18	Sequence 18, Appl
42	2.0	2137	2	US-08-828-451-18	Sequence 18, Appl
43	2.0	2140	2	US-08-541-033A-1	Sequence 1, Appli
44	2.0	2140	2	US-08-828-451-1	Sequence 1, Appli
45	2.0	2420	5	PCT-US93-00227-1	Sequence 1, Appli

Query Match 2.1%; Score 30; DB 4; Length 3300;  
Best Local Similarity 100.0%; Pred. No. 0.00092;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1387 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
DB 3238 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 3267

RESULT 2  
US-08-036-555B-134  
; Sequence 134, Application US/08036555B  
; Patent No. 5530109  
; GENERAL INFORMATION:  
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;  
; APPLICANT: Chennetti, Luisa; Waterfield, Michael; Marchioni, Mark;  
; APPLICANT: Chen, Mao Su; Hiles, Ian  
; TITLE OF INVENTION: Glial Mitogenic Factors, Their  
; TITLE OF INVENTION: Preparation and Use

NUMBER OF SEQUENCES: 184  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/036,555B  
FILING DATE: 24-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APRIL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.K. 91 07566.3  
FILING DATE: 10-APRIL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsai, Christine H.  
REGISTRATION NUMBER: 34,266  
REFERENCE/DOCKET NUMBER: LUD 5250.4  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1193  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-036-555B-134

Query Match 2.0%; Score 29; DB 1; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

## RESULT 3

US-08-469-569-134  
Sequence 134, Application US/08469569  
Patent No. 5606032  
GENERAL INFORMATION:  
APPLICANT: Goodearl, Andrew; Stroobant, Paul;  
APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;  
APPLICANT: Chen, Maio Su; Hiles, Ian  
TITLE OF INVENTION: Glial Mitogenic Factors, Their  
PREPARATION AND USE  
NUMBER OF SEQUENCES: 184  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA

ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,569  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/036,555  
FILING DATE: 24-MAR-1993  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APRIL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.K. 91 07566.3  
FILING DATE: 10-APRIL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsai, Christine H.  
REGISTRATION NUMBER: 34,266  
REFERENCE/DOCKET NUMBER: LUD 5250.4  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1193  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-469-569-134

Query Match 2.0%; Score 29; DB 1; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

## RESULT 4

US-08-249-322A-134  
Sequence 134, Application US/08249322A  
Patent No. 5716930  
GENERAL INFORMATION:  
APPLICANT: Goodearl, Andrew; Stroobant, Paul;  
APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;  
APPLICANT: Chen, Maio Su; Hiles, Ian  
TITLE OF INVENTION: Glial Mitogenic Factors, Their  
PREPARATION AND USE  
NUMBER OF SEQUENCES: 184  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS

```
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/249,322A
FILING DATE: 26-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/036,555
FILING DATE: 24-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,173
FILING DATE: 23-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/940,389
FILING DATE: 03-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/907,138
FILING DATE: 30-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/863,703
FILING DATE: 03-APRIL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.K. 91 07566.3
FILING DATE: 10-APRIL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Tsai, Christine H.
REGISTRATION NUMBER: 34,266
REFERENCE/DOCKET NUMBER: LUD 250.4
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 1193
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-249-322A-134

Query Match 2.0%; Score 29; DB 1; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
|||||
DB 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 5
US-08-469-526A-134
Sequence 134, Application US/08469526A
Patent No. 5792849
GENERAL INFORMATION:
APPLICANT: Goodearl, Andrew
APPLICANT: Stroobant, Paul
APPLICANT: Minghetti, Luisa
APPLICANT: Waterfield, Michael
APPLICANT: Marchionni, Mark
APPLICANT: Hiles, Ian
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
PREPARATION AND USE
NUMBER OF SEQUENCES: 187
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
```

```
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,526A
FILING DATE: 06 June 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/036,555
FILING DATE: 24-MAR-1993
APPLICATION NUMBER: 07/965,173
FILING DATE: 23-OCT-1992
APPLICATION NUMBER: 07/940,389
FILING DATE: 03-SEP-1992
APPLICATION NUMBER: 07/907,138
FILING DATE: 03-JUN-1992
APPLICATION NUMBER: 07/863,703
FILING DATE: 03-APRIL-1992
APPLICATION NUMBER: U.K. 91 07566.3
FILING DATE: 10-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bleker-Brady, Kristina
REGISTRATION NUMBER: 39,109
REFERENCE/DOCKET NUMBER: 04585/00200A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
INFORMATION FOR SEQ ID NO: 134:
SEQUENCE CHARACTERISTICS:
LENGTH: 1193
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-469-526A-134

Query Match 2.0%; Score 29; DB 1; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
|||||
DB 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 6
US-08-734-591A-134
Sequence 134, Application US/08734591A
Patent No. 5854220
GENERAL INFORMATION:
APPLICANT: Goodearl, Andrew
APPLICANT: Stroobant, Paul
APPLICANT: Minghetti, Luisa
APPLICANT: Waterfield, Michael
APPLICANT: Hiles, Ian
APPLICANT: Marchionni, Mark
APPLICANT: Chen, Mario
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
PREPARATION AND USE
NUMBER OF SEQUENCES: 187
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible Pentium
OPERATING SYSTEM: Windows95
SOFTWARE: WordPerfect (Version 7.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/734,591A
FILING DATE: 22-OCT-1996
CLASSIFICATION: 536
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/470,335
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 03-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 91 07566.3
; FILING DATE: 10-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 04585/00200P
; TELEPHONE: (617) 428-0200
; TELEFAX: (617) 428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-734-591A-134

Query Match 2.0%; Score 29; DB 2; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
|||||
DB 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 7
US-08-469-660-134
; Sequence 134, Application US/08469660
; Patent No. 5876973
; GENERAL INFORMATION:
; APPLICANT: Gwynne, David I.; Marchionni, Mark;
; APPLICANT: McBurney, Robert N.
; TITLE OF INVENTION: INHIBITORS OF CELL PROLIFERATION,
; TITLE OF INVENTION: THEIR PREPARATION AND USE
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; ZIP: 0211-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,660
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/011,396
```

```
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/984,085
; FILING DATE: 01-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/951,747
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/927,337
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04585/017004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: 200154
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-469-660-134

Query Match 2.0%; Score 29; DB 2; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
|||||
DB 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 8
US-08-341-018-3
; Sequence 3, Application US/08341018A
; Patent No. 6087323
; GENERAL INFORMATION:
; APPLICANT: Gwynne, David I.
; APPLICANT: Mahanthappa, Nagesh K.
; APPLICANT: Marchionni, Mark A.
; APPLICANT: Bermingham-McDonogh, Olivia
; APPLICANT: Goldin, Stanley M.
; APPLICANT: McBurney, Robert N.
; TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF
; TITLE OF INVENTION: CELLULAR COMMUNICATION
; FILE REFERENCE: 04585/041001
; CURRENT APPLICATION NUMBER: US/08/341,018A
; CURRENT FILING DATE: 1994-11-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1193
; TYPE: DNA
; ORGANISM: Bos taurus
; US-08-341-018-3

Query Match 2.0%; Score 29; DB 3; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
|||||
DB 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 9
US-08-470-335-134
; Sequence 134, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
```



APPLICANT: STROOBANT, PAUL  
APPLICANT: MINGHETTI, LUISA  
APPLICANT: WATERFIELD, MICHAEL  
APPLICANT: MARCHIONNI, MARK  
APPLICANT: CHEN, MARIO S.  
APPLICANT: HILES, IAN  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
PREPARATION AND USE  
FILE REFERENCE: 04585/00200B  
CURRENT APPLICATION NUMBER: US/08/470,335F  
CURRENT FILING DATE: 1995-06-06  
EARLIER APPLICATION NUMBER: 08/036,555  
EARLIER FILING DATE: 1993-03-24  
NUMBER OF SEQ ID NOS: 252  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 134  
LENGTH: 1193  
TYPE: DNA  
ORGANISM: Bos taurus  
US-08-470-335-134

Query Match 2.0%; Score 29; DB 3; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 10  
US-08-735-021-134  
Sequence 134, Application US/08735021B  
Patent No. 6194377  
GENERAL INFORMATION:  
APPLICANT: GOODEARL, ANDREW  
APPLICANT: STROOBANT, PAUL  
APPLICANT: MINGHETTI, LUISA  
APPLICANT: WATERFIELD, MICHAEL  
APPLICANT: MARCHIONNI, MARK  
APPLICANT: CHEN, MARIO S.  
APPLICANT: HILES, IAN  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
PREPARATION AND USE  
FILE REFERENCE: 04585/00200L  
CURRENT APPLICATION NUMBER: US/08/735,021B  
CURRENT FILING DATE: 1996-10-22  
EARLIER APPLICATION NUMBER: 08/472,065  
EARLIER FILING DATE: 1995-06-06  
EARLIER APPLICATION NUMBER: 08/036,555  
EARLIER FILING DATE: 1993-03-24  
EARLIER APPLICATION NUMBER: 07/965,173  
EARLIER FILING DATE: 1992-10-23  
EARLIER APPLICATION NUMBER: 07/940,389  
EARLIER FILING DATE: 1992-09-03  
EARLIER APPLICATION NUMBER: 07/907,138  
EARLIER FILING DATE: 1992-06-30  
EARLIER APPLICATION NUMBER: 07/863,703  
EARLIER FILING DATE: 1992-04-03  
NUMBER OF SEQ ID NOS: 192  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 134  
LENGTH: 1193  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (8)...(796)  
US-08-735-021-134

Query Match 2.0%; Score 29; DB 4; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 11  
US-08-734-664A-134  
Sequence 134, Application US/08734664A  
Patent No. 6204241  
GENERAL INFORMATION:  
APPLICANT: GOODEARL, ANDREW  
APPLICANT: STROOBANT, PAUL  
APPLICANT: MINGHETTI, LUISA  
APPLICANT: WATERFIELD, MICHAEL  
APPLICANT: MARCHIONNI, MARK  
APPLICANT: CHEN, MARIO  
APPLICANT: HILES, IAN  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
PREPARATION AND USE  
NUMBER OF SEQUENCES: 187  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Clark & Elbing LLP  
STREET: 176 Federal Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible Pentium  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/734,664A  
FILING DATE: 22-OCT-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/249,322  
FILING DATE: 26-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/036,555  
FILING DATE: 24-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: UK 91 07566.3  
FILING DATE: 10-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bleker-Brady, Kristina  
REGISTRATION NUMBER: 39,109  
REFERENCE/DOCKET NUMBER: 04585/00200J  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 428-0200  
TELEFAX: (617) 428-7045  
TELEX:  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1193  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-734-664A-134

Query Match 2.0%; Score 29; DB 4; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 12  
US-08-470-339-134  
; Sequence 134, Application US/08470339C  
; Patent No. 6232286  
; GENERAL INFORMATION:  
; APPLICANT: GOODEARL, ANDREW  
; APPLICANT: STROOBANT, PAUL  
; APPLICANT: MINGHETTI, LUISA  
; APPLICANT: WATERFIELD, MICHAEL  
; APPLICANT: MARCHIONNI, MARK  
; APPLICANT: CHEN, MARIO S.  
; APPLICANT: HILES, IAN  
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
; PREPARATION AND USE  
; FILE REFERENCE: 04585/002008  
; CURRENT APPLICATION NUMBER: US/08/470,339C  
; CURRENT FILING DATE: 1995-06-06  
; EARLIER APPLICATION NUMBER: 08/036,555  
; EARLIER FILING DATE: 1993-03-24  
; EARLIER APPLICATION NUMBER: 07/940,389  
; EARLIER FILING DATE: 1992-09-03  
; EARLIER APPLICATION NUMBER: 07/907,138  
; EARLIER FILING DATE: 1992-06-30  
; EARLIER APPLICATION NUMBER: 07/863,703  
; EARLIER FILING DATE: 1992-04-03  
; EARLIER APPLICATION NUMBER: 91 07566.3 GB  
; EARLIER FILING DATE: 1999-04-10  
; NUMBER OF SEQ ID NOS: 226  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 134  
; LENGTH: 1193  
; TYPE: DNA  
; ORGANISM: Bos taurus  
US-08-470-339-134

Query Match 2.0%; Score 29; DB 4; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 13  
US-08-467-602-134  
; Sequence 134, Application US/08467602C  
; Patent No. 6444642  
; GENERAL INFORMATION:  
; APPLICANT: Sklar, Robert  
; APPLICANT: Marchionni, Mark  
; APPLICANT: Gwynne, David I.  
; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND  
; DISORDERS  
; FILE REFERENCE: 04585/028003  
; CURRENT APPLICATION NUMBER: US/08/467,602C  
; CURRENT FILING DATE: 1995-06-06  
; EARLIER APPLICATION NUMBER: 08/209,204  
; EARLIER FILING DATE: 1994-03-08  
; EARLIER APPLICATION NUMBER: 08/059,022  
; EARLIER FILING DATE: 1993-05-06  
; NUMBER OF SEQ ID NOS: 420  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 134  
; LENGTH: 1193  
; TYPE: DNA  
; ORGANISM: Bos taurus  
US-08-467-602-134

Query Match 2.0%; Score 29; DB 4; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 14  
PCT-US94-05083C-130  
; Sequence 130, Application PC/TUS9405083C  
; GENERAL INFORMATION:  
; APPLICANT: Robert Sklar, Mark Marchionni,  
; APPLICANT: David I. Gwynne  
; TITLE OF INVENTION: METHODS FOR ALTERING  
; MUSCLE CONDITION  
; NUMBER OF SEQUENCES: 185  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360  
; MEDIUM TYPE: kb Storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/05083C  
; FILING DATE: 06-MAY-94  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/209,204  
; FILING DATE: 08-MAR-94  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/059,022  
; FILING DATE: 06-MAY-93  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 04585/028W01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 130:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1193  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
PCT-US94-05083C-130

Query Match 2.0%; Score 29; DB 5; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 15  
PCT-US95-06846A-134

```

; Sequence 134, Application PC/TUS9506846A
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew David; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06846A
; FILING DATE: 25-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/249,322
; FILING DATE: 26-MAY-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Norman D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5250.5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US95-06846A-134

```

```

Query Match 2.0%; Score 29; DB 5; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1388 AAAGTCACAAAAA...AAAAAAAAAAAA 1416
Db 1124 AAAGTCACAAAAA...AAAAAAAAAAAA 1152

```

```

Search completed: December 14, 2002, 17:04:10
Job time : 78 secs

```

THIS PAGE BLANK (USPTO)







```
; NAME/KEY: misc feature
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (47)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (52)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (60)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (76)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (149)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (164)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (172)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (187)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (190)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (198)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (204)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-11058
```

```
Query Match 2.3%; Score 32; DB 10; Length 206;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 79 GTAGGACCTACTCCCTATCCGTCGCGCGGG 110
|||||
Db 8 GTAGGACCTACTCCCTATCCGTCGCGCGGG 39
```

```
RESULT 7
US-09-867-701-5968/c
; Sequence 5968, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agiate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5968
; LENGTH: 311
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-867-701-5968
```

```
Query Match 2.0%; Score 29; DB 10; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
```

```
Db 40 AAGTCAAAAAAAAAAAAAAAAAAAAAA 12
|||||
```

```
RESULT 8
US-09-880-107-256/c
; Sequence 256, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 256
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AAL27741
; US-09-880-107-256
```

```
Query Match 2.0%; Score 29; DB 10; Length 564;
Best Local Similarity 100.0%; Pred. No. 0.0094;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
|||||
Db 43 AAGTCAAAAAAAAAAAAAAAAAAAAAA 15
```

```
RESULT 9
US-09-960-352-12666
; Sequence 12666, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12666
; LENGTH: 280
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 54-LIB3058-028-Q1-K1-F6
; US-09-960-352-12666
```

```
Query Match 2.0%; Score 28; DB 10; Length 280;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
|||||
Db 200 AAGTCAAAAAAAAAAAAAAAAAAAAAA 227
```

```
RESULT 10
US-09-867-701-2350/c
```



```
; Sequence 2350, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2350
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(319)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-2350

Query Match          2.0%; Score 28; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
      |||
Db 43 AAGTCAAAAAAAAAAAAAAAAAAAAAA 16

RESULT 11
US-09-770-445-359
; Sequence 359, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Fu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurlban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 359
; LENGTH: 941
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(941)
; OTHER INFORMATION: n = A,T,C or G
```

```
US-09-770-445-359
Query Match          2.0%; Score 28; DB 10; Length 941;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
      |||
Db 912 AAGTCAAAAAAAAAAAAAAAAAAAAAA 939

RESULT 12
US-09-925-301-142
; Sequence 142, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 142
; LENGTH: 1505
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1493)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1499)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1500)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-142

Query Match          2.0%; Score 28; DB 10; Length 1505;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
      |||
Db 1432 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1459

RESULT 13
US-09-874-628-5
; Sequence 5, Application US/09874628
; Patent No. US20020137133A1
; GENERAL INFORMATION:
; APPLICANT: WOZNEY, John
; APPLICANT: CELESTE, Anthony J.
; APPLICANT: THIES, R. Scott
; APPLICANT: YAMAJI, No. US20020137133A1oru
; TITLE OF INVENTION: RECEPTOR PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute Inc.- Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/874,628
; FILING DATE: 05-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,934
; FILING DATE: 17-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LAZAR, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5203
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 876 1170
; TELEFAX: 617 876 5851
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: CFK1-10a
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 474..2000
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-874-628-5

Query Match 2.0%; Score 28; DB 10; Length 3238;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 3155 AAGTCAAAAAAAAAAAAAAAAAAAAAA 3182

RESULT 14
US-09-777-745-1
; Sequence 1, Application US/09777745
; Patent No. US20010021702A1
; GENERAL INFORMATION:
; APPLICANT: CALMELS, THIERRY PAUL GERARD
; APPLICANT: SOUCHET, MICHEL LOUIS
; APPLICANT: LEGER, ISABELLE MARIE
; APPLICANT: TOURTELIER, LAURENCE NATHALIE PATRICIA
; APPLICANT: BRIL, ANTOINE MICHEL ALAIN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30321-C1
; CURRENT APPLICATION NUMBER: US/09/777,745
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: EP 98400014.1
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1997-05-27
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3592
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-777-745-1

Query Match 2.0%; Score 28; DB 10; Length 3592;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 3155 AAGTCAAAAAAAAAAAAAAAAAAAAAA 3182

RESULT 14
US-09-777-745-1
; Sequence 1, Application US/09777745
; Patent No. US20010021702A1
; GENERAL INFORMATION:
; APPLICANT: CALMELS, THIERRY PAUL GERARD
; APPLICANT: SOUCHET, MICHEL LOUIS
; APPLICANT: LEGER, ISABELLE MARIE
; APPLICANT: TOURTELIER, LAURENCE NATHALIE PATRICIA
; APPLICANT: BRIL, ANTOINE MICHEL ALAIN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30321-C1
; CURRENT APPLICATION NUMBER: US/09/777,745
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: EP 98400014.1
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: EP 98400015.8
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: UK 9710910.2
; PRIOR FILING DATE: 1997-05-27
; PRIOR APPLICATION NUMBER: 09/082,271
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3592
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-777-745-1

Query Match 2.0%; Score 28; DB 10; Length 3592;
Best Local Similarity 100.0%; Pred. No. 0.02;
```

```

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 3559 AAGTCAAAAAAAAAAAAAAAAAAAAAA 3586

RESULT 15
US-09-777-745-5
; Sequence 5, Application US/09777745
; Patent No. US20010021702A1
; GENERAL INFORMATION:
; APPLICANT: CALMELS, THIERRY PAUL GERARD
; APPLICANT: SOUCHET, MICHEL LOUIS
; APPLICANT: LEGER, ISABELLE MARIE
; APPLICANT: TOURTELIER, LAURENCE NATHALIE PATRICIA
; APPLICANT: BRIL, ANTOINE MICHEL ALAIN
; APPLICANT: DOE, TRUDY RACHEL
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30321-C1
; CURRENT APPLICATION NUMBER: US/09/777,745
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: EP 98400014.1
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: EP 98400015.8
; PRIOR FILING DATE: 1998-01-07
; PRIOR FILING DATE: 1997-05-27
; PRIOR APPLICATION NUMBER: 09/082,271
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3627
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-777-745-5

Query Match 2.0%; Score 28; DB 10; Length 3627;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 3594 AAGTCAAAAAAAAAAAAAAAAAAAAAA 3621

Search completed: December 14, 2002, 17:56:17
Job time : 102 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2002, 16:58:17 ; Search time 2016 Seconds  
(without alignments)  
11375.401 Million cell updates/sec

Title: US-09-357-675C-1  
Perfect score: 1416  
Sequence: 1 gccactcgtcgcgctnt.....aaaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC  
Gap 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	844	59.6	844	9	AL529152
2	813	57.4	890	9	AL520768
3	798	56.4	943	14	BM804704
4	691	48.8	960	9	AL522373
5	677	47.8	782	14	BQ441752
6	675	47.7	1048	14	BQ059007

C 7	660	45.6	677	14	BQ581760
C 8	645	45.6	659	13	BM666784
C 9	625	44.1	625	13	BI711300
C 10	610	43.1	729	14	BM975958
C 11	606	42.8	613	14	BM690873
C 12	604	42.7	846	13	BI769604
C 13	589	41.6	750	14	BM979227
C 14	585	41.3	992	9	AL520767
C 15	584	41.2	586	9	AI797259
C 16	577	40.7	600	14	BM726885
C 17	573	40.5	1150	14	BM925975
C 18	572	40.4	579	13	BI711712
C 19	571	40.3	736	9	AI668782
C 20	563	39.8	624	14	BQ582098
C 21	557	39.3	559	13	BM141822
C 22	548	38.7	550	13	BM141736
C 23	547	38.6	766	14	BM679998
C 24	545	38.5	793	12	BG436916
C 25	543	38.3	746	13	BI752623
C 26	539	38.1	938	14	BQ073413
C 27	534	37.7	764	12	BG762506
C 28	524	37.0	550	9	AI017543
C 29	510	36.0	710	9	AI797380
C 30	510	36.0	781	12	BG532265
C 31	510	36.0	928	13	BI822844
C 32	505	35.7	537	13	BM142010
C 33	503	35.5	552	13	BM142111
C 34	487	34.4	560	10	AW182514
C 35	468	33.1	576	10	AW956706
C 36	468	33.1	657	13	BI916460
C 37	461	32.6	461	9	AA907376
C 38	459	32.4	946	13	BI757823
C 39	457	32.3	790	12	BG703176
C 40	452	31.9	813	13	BI755493
C 41	451	31.9	633	14	BQ002524
C 42	448	31.6	791	12	BG776560
C 43	446	31.5	579	9	AI559188
C 44	444	31.4	749	10	AW073366
C 45	434	30.6	482	9	AI085503

#### ALIGNMENTS

RESULT 1  
AL529152  
LOCUS AL529152 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5  
DEFINITION prime, mRNA  
ACCESSION AL529152  
VERSION AL529152.1 GI:12792645  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 844)  
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Pollayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
Location/Qualifiers  
Source

1..844  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DD004YM06"  
/clone\_lib="LTI\_NFL001\_NBC4"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"



Db 370 CACTGGGTGGGAACCTTTGGGAAGAAATACACCCAGCTTCCAGGGAATGTGACTCTGCG 429  
QY 478 TGTCTCTGGGTGGTTTCCATGATGCGGTGGCAAGACTGGGAGCAGACTCAGAAAATCTACA 537  
Db 430 TGTCTCTGGGTGGTTTCCATGATGCGGTGGCAAGACTGGGAGCAGACTCAGAAAATCTACA 489  
QY 538 ATTGTACATGCTGTGCTGAACACAGCAAGGGGCAAGTGTGCTGCTGCTGCTGCTGCTGCT 597  
Db 490 ATTGTACATGCTGTGCTGAACACAGCAAGGGGCAAGTGTGCTGCTGCTGCTGCTGCTGCT 549  
QY 598 TGTGTGACCTAGATGCTGCAAGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 657  
Db 550 TGTGTGACCTAGATGCTGCAAGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 609  
QY 658 GGGCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717  
Db 610 GGGCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669  
QY 718 ATGACATGCGGTTCCTGAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 777  
Db 670 ATGACATGCGGTTCCTGAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 729  
QY 778 CTTATCTTCACTTTTGTGATCATACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 837  
Db 730 CTTATCTTCACTTTTGTGATCATACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 789  
QY 838 CCGTGTCTATCAAGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 897  
Db 790 CCGTGTCTATCAAGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 849  
QY 898 AGAAGAGCAAGTTATGCGCCACAGCATGGTGG 930  
Db 850 AGAAGAGCAAGTTATGCGCCACAGCATGGTGG 882

## RESULT 3

BM804704  
LOCUS  
DEFINITION  
5', mRNA sequence.  
AGENCY: 6494868 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:5725375  
EST. 5725375

BM804704  
BM804704.1 GI:19121527

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 943)

NIH-MGC <http://mgc.ncl.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: Invitrogen

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM12716 row: c column: 08

High quality sequence stop: 716.

Location/Qualifiers

1..943

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5725375"

/lab\_host="DH10B"

/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;

Site: 1: EcorV (destroyed); Site: 2: NotI; RNA source pool

of three ovaries, from females ranging in age from 38 to

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

49 yo. Library is oligo-dT primed and directionally cloned

(EcorV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code "036."

BASE COUNT 198 a 270 c 215 t  
ORIGIN

Query Match 56.4%; Score 798; DB 14; Length 943;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GCTGGGCTTATCACCAGGCTCCTCACAGATTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 169  
Db 54 GCTGGGCTTATCACCAGGCTCCTCACAGATTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 113  
QY 170 GATACCTCAACTCTCAGTACTTTGTCTCAGCCAGGCTCAGGCTCAGGCTCAGGCTCAGGCT 229  
Db 114 GATACCTCAACTCTCAGTACTTTGTCTCAGCCAGGCTCAGGCTCAGGCTCAGGCTCAGGCT 173  
QY 230 TTCTCTCTCGGAACCTGCCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 289  
Db 174 TTCTCTCTCGGAACCTGCCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 233  
QY 290 ACAGAACTTTAAACATGT 349  
Db 234 ACAGAACTTTAAACATGT 293  
QY 350 GCTTTCTCTGCTGAGGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 409  
Db 294 GCTTTCTCTGCTGAGGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 353  
QY 410 GTCTGAACCACTGGGTGGGAACTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 469  
Db 354 GTCTGAACCACTGGGTGGGAACTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 413  
QY 470 ACTCTGGCTGTCTCTGGGTGGTTCATCAGGCTGGCCAGGCTGGGAGCAGACTCAGAA 529  
Db 414 ACTCTGGCTGTCTCTGGGTGGTTCATCAGGCTGGCCAGGCTGGGAGCAGACTCAGAA 473  
QY 530 AATCTACATTTGTACGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 589  
Db 474 AATCTACATTTGTACGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 533  
QY 590 GACATCTGT 649  
Db 534 GACATCTGT 593  
QY 650 CATGCTGGGCTGCTTTGT 709  
Db 594 CATGCTGGGCTGCTTTGT 653  
QY 710 TGTCTGT 769  
Db 654 TGTCTGT 713  
QY 770 GATCTTACCTTCT 829  
Db 714 GATCTTACCTTCT 773  
QY 830 GCTGGGGGCTGT 889  
Db 774 GCTGGGGGCTGT 833  
QY 890 CCACCATGAGAGAGAGC 907  
Db 834 CCACCATGAGAGAGAGC 851

## RESULT 4

AL522373/c

LOCUS

AL522373

DEFINITION

AL522373 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB008YM02 3

960 bp mRNA linear EST 13-FEB-2001

CS0DB008YM02 3

```
prime, mRNA sequence.
AL522373
VERSION AL522373.1 GI:12785866
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB08YM02"
/clone_lib="LRI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 225 a 259 c 255 g 217 t 4 others
ORIGIN
Query Match 48.8%; Score 691; DB 9; Length 960;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 911; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 438 GAAGATAACCCAGCTTCCAGGGAATGTGACTCTGGCTCTCCTTGGGTGGTTTCCAT 497
Db 922 GAAGATAACCCAGCTTCCAGGGAATGTGACTCTGGCTCTCCTTGGGTGGTTTCCAT 863
QY 498 GAGCGTGGCCACAGACTGGGAGCAGACTCAGAAATCTACAATTGTACGTGCTGTGAC 557
Db 862 GAGCGTGGCCACAGACTGGGAGCAGACTCAGAAATCTACAATTGTACGTGCTGTGAC 803
QY 558 AGCAAGGGCCAGTAGTGCCACTTACAGAGACACATCTGTGACGTAGAGATTCCA 617
Db 802 AGCAAGGGCCAGTAGTGCCACTTACAGAGACACATCTGTGACGTAGAGATTCCA 743
QY 618 GGGCAGGGCCCTATGTGTGAAGCAACTTACATGCCCTGGGCCCGAGCTTGGAGTCACT 677
Db 742 GGGCAGGGCCCTATGTGTGAAGCAACTTACATGCCCTGGGCCCGAGCTTGGAGTCACT 683
QY 678 GTACGACACACAGCAGGCAAGATTGGTCTAGCTGTCTGTATGACATCGGTTCCCTGAA 737
Db 682 GTACGACACACAGCAGGCAAGATTGGTCTAGCTGTCTGTATGACATCG-GTTCCTGAA 624
QY 738 CTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGA 797
Db 623 CTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGA 564
QY 798 TCATTACAGGCCAGCCACCTGGGAGGTGTGTGCGGGGCCCGCTGCTATCCAAACCCAG 857
Db 563 TCATTACAGGCCAGCCACCTGGGAGGTGTGTGCGGGGCCCGCTGCTATCCAAACCCAG 504
QY 858 TGTATGTAGTGGCAGCAGCAGTGTGACGCCACCATGACAAGAGCAAGTATATGCG 917
Db 503 TGTATGTAGTGGCAGCAGCAGTGTGACGCCACCATGACAAGAGCAAGTATATGCG 444
```

```
QY 918 CACAGCATGGTGTAGACCCCTGGGAAACAGTGGTGGCCGCTGCTCTGAGGGCCAGCG 977
Db 443 CACAGCATGGTGTAGACCCCTGGGAAACAGTGGTGGCCGCTGCTCTGAGGGCCAGCG 384
QY 978 CTCTGCCCTGGCCGAATAGACCTCAACTATCTGCGACAGTTGGCCCGACACCTGCTGTG 1037
Db 383 CTCTGCCCTGGCCGAATAGACCTCAACTATCTGCGACAGTTGGCCCGACACCTGCTGTG 324
QY 1038 TTCCAGCAGCCGAGCCCTGACCTCTATGCAATCTGGGTACCCCACTGCTTTAAGACTTG 1097
Db 323 TTCCAGCAGCCGAGCCCTGACCTCTATGCAATCTGGGTACCCCACTGCTTTAAGACTTG 264
QY 1098 ACTTCTGTGAGTTTAGACCTGCCCTCCACACCCACCCCTGCCACTATGAGCTAGTGCTC 1157
Db 263 ACTTCTGTGAGTTTAGACCTGCCCTCCACACCCACCCCTGCCACTATGAGCTAGTGCTC 204
QY 1158 ATGTGACTGGAGGAGGATCCAGGCACAGCTCCCTCCTCACTTGGAGACCTTGACTCTCT 1217
Db 203 ATGTGACTGGAGGAGGATCCAGGCACAGCTCCCTCCTCACTTGGAGACCTTGACTCTCT 144
QY 1218 TGATGSAACACAGATGGCTGCTTGGGAAAGAACTTTCACTGAGCTTCACTGAGGTC 1277
Db 143 TGATGSAACACAGATGGCTGCTTGGGAAAGAACTTTCACTGAGCTTCACTGAGGTC 84
QY 1278 AGACTGCAGTTTCAGAAAGGTGGAATTTATATAGTCACTTGTATTTCATGGAACCTGA 1337
Db 83 AGACTGCAGTTTCAGAAAGGTGGAATTTATATAGTCACTTGTATTTCATGGAACCTGA 24
QY 1338 AGTTCTGCTGAGGG 1351
Db 23 AGTTCTGCTGAGGG 10
RESULT 5
BO441752
LOCUS BO441752
DEFINITION BO441752 782 bp mRNA linear EST 24-MAY-2002
5' mRNA sequence.
ACCESSION BO441752
VERSION BO441752.1 GI:21180828
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 782)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LIAW13223 row: c column: 13
High quality sequence stop: 672.
Location/Qualifiers
1..782
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6019908"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site:1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
```

BASE COUNT	185 a	209 c	200 g	186 t	2 others
ORIGIN					
Query Match	47.8%; Score 677; DB 14; Length 782;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 677; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	696	AAGATTGGTCTAGCTGCTGCTATGACATGCGGTTCCTCGAATCTCTCTGGCATTTGGCT	755		
Db	1	AAGATTGGTCTAGCTGCTGCTATGACATGCGGTTCCTCGAATCTCTCTGGCATTTGGCT	60		
QY	756	CAAGCTGGACGACAGATACCTTACCTTCTCAGCTTTTGGATCCATTACAGGCCAGCC	815		
Db	61	CAAGCTGGACGACAGATACCTTACCTTCTCAGCTTTTGGATCCATTACAGGCCAGCC	120		
QY	816	CACCTGGAGGTGTGCTGGCGGCCCTGTCTATCGAAACCCAGTGTATGTAGTGGCAGCA	875		
Db	121	CACCTGGAGGTGTGCTGGCGGCCCTGTCTATCGAAACCCAGTGTATGTAGTGGCAGCA	180		
QY	876	GCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGTTGGTAGAC	935		
Db	181	GCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGTTGGTAGAC	240		
QY	936	CCCTGGGGAACAGTGTGGCCCCCTGCTCTGAGGGGCCAGGCTCTGCCTTGCCCGAATA	995		
Db	241	CCCTGGGGAACAGTGTGGCCCCCTGCTCTGAGGGGCCAGGCTCTGCCTTGCCCGAATA	300		
QY	996	GACCTCAACTATCTGGACAGTTGGCGCCGACACCTGCTGTGTCCAGCACCGCAGGCCT	1055		
Db	301	GACCTCAACTATCTGGACAGTTGGCGCCGACACCTGCTGTGTCCAGCACCGCAGGCCT	360		
QY	1056	GACCTCTATGGCAATCTGGGTACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGAC	1115		
Db	361	GACCTCTATGGCAATCTGGGTACCCACTGTCTTAAGACTTGACTTCTGTGAGTTTAGAC	420		
QY	1116	CTGCCCTCCACCCACCCTGCCACTATGACTAGTGTCTCATGTGACTTTGAGGCAGG	1175		
Db	421	CTGCCCTCCACCCACCCTGCCACTATGACTAGTGTCTCATGTGACTTTGAGGCAGG	480		
QY	1176	ATCCAGGCACAGCTCCCTCTCAGTTGGAGAACCTTGACTCTCTTGATGGAAACACAGATGG	1235		
Db	481	ATCCAGGCACAGCTCCCTCTCAGTTGGAGAACCTTGACTCTCTTGATGGAAACACAGATGG	540		
QY	1236	CTGCTTGGGAAAGAACTTTCACCTGAGCTTCACCTGAGTCAAGCTGCACTTTTCAGAAA	1295		
Db	541	CTGCTTGGGAAAGAACTTTCACCTGAGCTTCACCTGAGTCAAGCTGCACTTTTCAGAAA	600		
QY	1296	GGTGGAAATTTATATAGTCAATTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGGCTGA	1355		
Db	601	GGTGGAAATTTATATAGTCAATTTTATTTTCATGGAACCTGAAGTTCTGCTGAGGGCTGA	660		
QY	1356	GCAGCACTGGCAATTGAA	1372		
Db	661	GCAGCACTGGCAATTGAA	677		
RESULT 6					
BQ059007					
LOCUS					
DEFINITION	BQ059007 1048 bp mRNA linear EST 29-MAR-2002				
ACCESSION	AGENCOURT_6808192 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5814981				
VERSION	5', mRNA sequence.				
KEYWORDS	BQ059007.1 GI:19818347				
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
JOURNAL	1 (bases 1 to 1048)				
COMMENT	NIH-MGC http://mgi.nci.nih.gov/				
	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				

BASE COUNT	216 a	304 c	289 g	239 t	
ORIGIN					
Query Match	47.7%; Score 675; DB 14; Length 1048;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 675; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	110	GCTGGGCTTCATCACCAGGCGCTCCTCACAGATTCTCTGTCCTCTCTGTGCTGGACTCCG	169		
Db	65	GCTGGGCTTCATCACCAGGCGCTCCTCACAGATTCTCTGTCCTCTCTGTGCTGGACTCCG	124		
QY	170	GATACCTCAACTCTCAGTACTTTGTGTCAGCCAGCCAGAGCCATGGCTATCTCCTC	229		
Db	125	GATACCTCAACTCTCAGTACTTTGTGCTAGCCAGCCAGAGCCATGGCTATCTCCTC	184		
QY	230	TTCTCTCTCGGAATGCCCTGTGCTGCTGTGCCAGGTAAACATGACGCCAGACAAGCA	289		
Db	185	TTCTCTCTCGGAATGCCCTGTGCTGCTGTGCCAGGTAAACATGACGCCAGACAAGCA	244		
QY	290	ACAGAACTTTAAACATGCTGAGCTGTTTCGAGAGGCTGCCAGACTGGGTGCCCT	349		
Db	245	ACAGAACTTTAAACATGCTGAGCTGTTTCGAGAGGCTGCCAGACTGGGTGCCCT	304		
QY	350	GGCTTTCTGCGCTGAGGCATTTGACTTTCATTGCACGGGACCCCTGCAGAGACGCTACACCT	409		
Db	305	GGCTTTCTGCGCTGAGGCATTTGACTTTCATTGCACGGGACCCCTGCAGAGACGCTACACCT	364		
QY	410	GTCTGAACCACTGGGTGGGAAACTTTTGGAAAGTAATACACCCAGCTTCCCAGGAATGTGG	469		
Db	365	GTCTGAACCACTGGGTGGGAAACTTTTGGAAAGTAATACACCCAGCTTCCCAGGAATGTGG	424		
QY	470	ACTCTGGCTGCTCCTTGGGTGGTTTCCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAA	529		
Db	425	ACTCTGGCTGCTCCTTGGGTGGTTTCCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAA	484		
QY	530	AATCTACAATTTGTCAGTGTCTGAACAGCAAGGGGCGAGTAGTGGCCACTTACAGGAA	589		
Db	485	AATCTACAATTTGTCAGTGTCTGAACAGCAAGGGGCGAGTAGTGGCCACTTACAGGAA	544		
QY	590	GACACATCTGTGTGACCTAGAGATTCCAGGGCAGGGGCCCTTATGTGTGAAAGCAACTCTAC	649		
Db	545	GACACATCTGTGTGACCTAGAGATTCCAGGGCAGGGGCCCTTATGTGTGAAAGCAACTCTAC	604		
QY	650	CATGCTGGGCCCGCTCTGTGAGTCACTGTGTCAGCACACACAGCAGGCAAGATTGGTCTAGC	709		
Db	605	CATGCTGGGCCCGCTCTGTGAGTCACTGTGTCAGCACACACAGCAGGCAAGATTGGTCTAGC	664		

Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Lou Staudt  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2069 row: h column: 22  
High quality sequence start: 725.  
Location/Qualifiers  
1..1048  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5814981"  
/clone\_lib="NIH\_MGC\_99"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: pOTB7; Site1: XhoI; Site2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source

```

QY 710 TGCTGCTATGACATGGGTTCCCTGAACCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 769
|||||
Db 665 TGCTGCTATGACATGGGTTCCCTGAACCTCTCTGGCATTGGCTCAAGCTGGAGCAGA 724
|||||
QY 770 GATACTTACCTATCC 784
|||||
Db 725 GATACTTACCTATCC 739
|||||

RESULT 7
BQ581760/c 677 bp mRNA linear EST 20-JUN-2002
LOCUS 1112906.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029771
DEFINITION 3' similar to TR:076091 O76091 NITRILASE HOMOLOG 1.; mRNA
sequence.
ACCESSION BQ581760
VERSION BQ581760.1 GI:21494649
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 677)
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 483.
FEATURES
Location/Qualifiers
1..677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:602971"
/clone_lib="Human insulinoma"
/tissue_type="Insulinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
BASE COUNT 158 a 170 c 179 g 170 t
ORIGIN

Query Match 46.6%; Score 660; DB 14; Length 677;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 CCTGAACCTCTCTGTCATTGGCTCAAGCTGGAGCAGAGATACCTTACCTTACCTTACGCT 791
|||||

```

```

Db 677 CCTGAACCTCTCTGTCATTGGCTCAAGCTGGAGCAGAGATACCTTACCTTACGCT 618
|||||
QY 792 TTTGGATCATTACAGGCCCGCCAGCCACACTGGGAGGTGTTGCTCGGGCCCGTCTATCGAA 851
|||||
Db 617 TTTGGATCATTACAGGCCCGCCAGCCACACTGGGAGGTGTTGCTCGGGCCCGTCTATCGAA 558
|||||
QY 852 ACCCAGTGTCTATGTAGTGGCAGCAGACAGTGTGGAGCCACCATGAGAGAGCAAGT 911
|||||
Db 557 ACCCAGTGTCTATGTAGTGGCAGCAGACAGTGTGGAGCCACCATGAGAGAGCAAGT 498
|||||
QY 912 TATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGTCTCTGAGGGG 971
|||||
Db 497 TATGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGTCTCTGAGGGG 438
|||||
QY 972 CCAGGCTCTGCTGGTGGCCGAATAGACCTCAACTATCTCGGACAGTGGCGCGACACCTG 1031
|||||
Db 437 CCAGGCTCTGCTGGTGGCCGAATAGACCTCAACTATCTCGGACAGTGGCGCGACACCTG 378
|||||
QY 1032 CCGTGTGTTCCAGCAGCGAGGCTGACCTCTATGCAATCTGGGTCAACCACTGTCTTAA 1091
|||||
Db 377 CCGTGTGTTCCAGCAGCGAGGCTGACCTCTATGCAATCTGGGTCAACCACTGTCTTAA 318
|||||
QY 1092 GACTTGACTTCTGTGAGTTTACACCTGCCCCCTCCACACCCACCTGCCACTATGAGCTA 1151
|||||
Db 317 GACTTGACTTCTGTGAGTTTACACCTGCCCCCTCCACACCCACCTGCCACTATGAGCTA 258
|||||
QY 1152 GTGCTCATGTACTGTGGAGGAGGAGTCCAGGCACAGTCCCTCCTCATCTGGAGAACTTGA 1211
|||||
Db 257 GTGCTCATGTACTGTGGAGGAGGAGTCCAGGCACAGTCCCTCCTCATCTGGAGAACTTGA 198
|||||
QY 1212 CTTCTTGATGAACACAGATGGGCTGCTGTGGGAAGAACTTTCACCTGAGCTTCACT 1271
|||||
Db 197 CTTCTTGATGAACACAGATGGGCTGCTGTGGGAAGAACTTTCACCTGAGCTTCACT 138
|||||
QY 1272 GAGGTCAGACTGCAGTTTCAGAAAGTGGAATTTTATATAGTCATTGTTTATTTTCATGA 1331
|||||
Db 137 GAGGTCAGACTGCAGTTTCAGAAAGTGGAATTTTATATAGTCATTGTTTATTTTCATGA 78
|||||
QY 1332 AACTGAAGTTCTGCTGAGGGCTGAGCAGCAGTGGGATGAAAATATATAATATATAAAG 1391
|||||
Db 77 AACTGAAGTTCTGCTGAGGGCTGAGCAGCAGTGGGATGAAAATATATAATATATAAAG 18
|||||

RESULT 8
BQ666784/c 659 bp mRNA linear EST 27-FEB-2002
LOCUS UI-E-CL1-aez-a-24-0-UI.s1 UI-E-CL1 Homo sapiens cDNA clone
DEFINITION UI-E-CL1-aez-a-24-0-UI 3', mRNA sequence.
ACCESSION BQ666784
VERSION BQ666784.1 GI:18974615
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 659)
REFERENCE Bonaldo,M.F., Lemon,G. and Soares,M.B.
AUTHORS Normalization and subtraction: two approaches to facilitate gene
TITLE discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

```



Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 1-35, >AT-rich#Low\_complexity  
 Seq primer: M13 Forward  
 POLYA=Yes.

## FEATURES

source  
 1. .859  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-CL1-aez-a-24-0-UI"  
 /clone\_lib="UI-E-CL1"  
 /tissue\_type="human retina"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; UI-E-CL1 is a normalized cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
 TAG\_Lib=UI-E-CL1  
 TAG\_RIS08=human retina  
 TAG\_SEQ=CCGCG"  
 BASE COUNT 153 a 168 c 173 g 165 t  
 ORIGIN

Query Match 45.6%; Score 645; DB 13; Length 659;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 747 GCATTGGCTCAAGCTGAGCAGAGATACCTACCTATCTTACGCTTTTGATCCATTACA 806  
 Db 659 GCATTGGCTCAAGCTGAGCAGAGATACCTATCTTACGCTTTTGATCCATTACA 600  
 QY 807 GCGCCAGCCCACTGGAGGTGTTGCTCGCGGCCGCTGCTATCGAAACCCAGCTGCTATGTA 866  
 Db 599 GCGCCAGCCCACTGGAGGTGTTGCTCGCGGCCGCTGCTATCGAAACCCAGCTGCTATGTA 540  
 QY 867 GTGGCAGCAGCAGTGTGACCCCACTATGAGAGAGAGCAAGTTATGCCACACGATG 926  
 Db 539 GTGGCAGCAGCAGTGTGACCCCACTATGAGAGAGAGCAAGTTATGCCACACGATG 480  
 QY 927 GTGGTAGACCCCTGGGAAACAGTGTGGCCGCTGCTGAGGGCCAGGCCCTCGCCCT 986  
 Db 479 GTGGTAGACCCCTGGGAAACAGTGTGGCCGCTGCTGAGGGCCAGGCCCTCGCCCT 420  
 QY 987 GCGCGAATAGACCTCAACTATCTGCGACAGTGTGCGCGACACCTGCTGTGTCCAGCAC 1046  
 Db 419 GCGCGAATAGACCTCAACTATCTGCGACAGTGTGCGCGACACCTGCTGTGTCCAGCAC 360  
 QY 1047 CGCAGCGCTGACCTCTATGGCAATCTGGTGCACCCACTGCTTTAAGACCTGACTTCTGTG 1106  
 Db 359 CGCAGCGCTGACCTCTATGGCAATCTGGTGCACCCACTGCTTTAAGACCTGACTTCTGTG 300  
 QY 1107 AGTTTAGACCTGCTCCACCCACCCCTGCCACTATGAGTACTGCTCATGTGACTT 1166  
 Db 299 AGTTTAGACCTGCTCCACCCACCCCTGCCACTATGAGTACTGCTCATGTGACTT 240  
 QY 1167 GGAGGAGGATCCAGGCACAGCTCCCTCCTACCTTGGAGAACCTTGACTCTCTTATGATGAAC 1226  
 Db 239 GGAGGAGGATCCAGGCACAGCTCCCTCCTACCTTGGAGAACCTTGACTCTCTTATGATGAAC 180  
 QY 1227 ACAGATGGGCTGCTGGGAAAGAAACTTTACCTGAGCTTCACCTGAGCTCAGACTGCAG 1286

Db 179 ACAGATGGGCTGCTGGGAAAGAAACTTTACCTGAGCTTCACCTGAGCTCAGACTGCAG 120  
 QY 1287 TTTTCAGAAAGGTGGAAATTTTATATAGTCTATTGTTTATTTTCATGGAACCTGAAGTTCTGCT 1346  
 Db 119 TTTTCAGAAAGGTGGAAATTTTATATAGTCTATTGTTTATTTTCATGGAACCTGAAGTTCTGCT 60  
 QY 1347 GAGGGCTGAGCAGCACTGGCATTGAAAAATAATAATCATAAAG 1391  
 Db 59 GAGGGCTGAGCAGCACTGGCATTGAAAAATAATAATCATAAAG 15  
 RESULT 9  
 B7171300/c  
 LOCUS  
 DEFINITION  
 B7171300 625 bp mRNA linear EST 11-MAR-2002  
 id96ff12.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5023654  
 3' similar to TR:076091 076091 NITRILASE HOMOLOG 1. ; mRNA  
 sequence.  
 B7171300  
 EST.  
 B7171300.1 GI:15686995  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 625)  
 REFERENCE  
 AUTHORS  
 Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
 Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
 Hillier,L., Narra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A.,  
 Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas  
 , M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,  
 Jackson,Y. and Bowers,Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. J. Ferrer In vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
 University Genome Sequencing Center For information on obtaining a  
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
 Seq primer: 40UP from Gibco  
 High quality sequence stop: 481.  
 Location/Qualifiers  
 1. .625  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5023654"  
 /clone\_lib="Human insulinoma"  
 /tissue\_type="Insulinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site.1:  
 XhoI; Site.2: EcoRI; Constructed with lambda ZAPII system  
 (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue following the Washington  
 University protocol  
 (http://genome.wustl.edu/est/lambda\_protocol.shtml).  
 Please contact Hiroshi Inoue, MD/PhD for further  
 information on this library (Metabolism Division, Permutt  
 Laboratory, Washington University School of Medicine, Box  
 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: This  
 is a Washington University Pancreas EST project library."  
 BASE COUNT 139 a 183 c 161 g 142 t  
 ORIGIN  
 Query Match 44.1%; Score 625; DB 13; Length 625;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 333 AGACTGGTGCCTGCTGCTTCCCTGCTGAGGCATTTGACTTCAATTGCAAGGACCT 392
|||||
Db 625 AGACTGGTGCCTGCTGCTTCCCTGCTGAGGCATTTGACTTCAATTGCAAGGACCT 566
|||||
QY 393 GCAGAGACGCTACACCTGTCTGAACACACTGGGTGGGAACTTTTGGGAAGATACACCCAG 452
|||||
Db 565 GCAGAGACGCTACACCTGTCTGAACACACTGGGTGGGAACTTTTGGGAAGATACACCCAG 506
|||||
QY 453 CTTGCCAGGGAATGTGGACTCTGGCTGCTTGGGTGGTTCATGAGCGTGGCCAAAGAC 512
|||||
Db 505 CTTGCCAGGGAATGTGGACTCTGGCTGCTTGGGTGGTTCATGAGCGTGGCCAAAGAC 446
|||||
QY 513 TGGGACGAGACTCAGAAAATCTACAATTTGCTGACGTCTGCTGAACAGCAAAAGGGCAGTA 572
|||||
Db 445 TGGGACGAGACTCAGAAAATCTACAATTTGCTGACGTCTGCTGAACAGCAAAAGGGCAGTA 386
|||||
QY 573 GTGGCCACTTACAGGAAGACACATCTGTGTGAGATTCAGAGATTCAGGGGAGGGCCTATG 632
|||||
Db 385 GTGGCCACTTACAGGAAGACACATCTGTGTGAGATTCAGAGATTCAGGGGAGGGCCTATG 326
|||||
QY 633 TGTGAAGCAACTCTACCATGCTGGGCCAGCTTGTGAGTCACCTGTGACGACACACCA 592
|||||
Db 325 TGTGAAGCAACTCTACCATGCTGGGCCAGCTTGTGAGTCACCTGTGACGACACACCA 266
|||||
QY 693 GGCAGATTGGTCTAGCTGTCTGTATGACATGCGGCTTCCCTGAACTCTCTCTGGCATG 752
|||||
Db 265 GGCAGATTGGTCTAGCTGTCTGTATGACATGCGGCTTCCCTGAACTCTCTCTGGCATG 206
|||||
QY 753 GCTCAAGCTGGAGCAGAGATTAACCTATCCTTCAGCTTTTGGATTCATACAGGCCCA 812
|||||
Db 205 GCTCAAGCTGGAGCAGAGATTAACCTATCCTTCAGCTTTTGGATTCATACAGGCCCA 146
|||||
QY 813 GCCCACTGGGAGGTGTGTGCGGGCCGCTGCTATGCAACCCAGTGTATGATGGCA 872
|||||
Db 145 GCCCACTGGGAGGTGTGTGCGGGCCGCTGCTATGCAACCCAGTGTATGATGGCA 86
|||||
QY 873 GCAGCAGCTGTGGAGCCCAACCATGAGAGAGCAAGTATATGGCCACAGCATGTGGTA 932
|||||
Db 85 GCAGCAGCTGTGGAGCCCAACCATGAGAGAGCAAGTATATGGCCACAGCATGTGGTA 26
|||||
QY 933 GACCCCTGGGAAACAGTGTGGGCC 957
|||||
Db 25 GACCCCTGGGAAACAGTGTGGGCC 1
|||||

RESULT 10
BM975958/c
LOCUS
DEFINITION
  729 bp mRNA linear EST 21-MAR-2002
  UI-CF-EN1-acw-1-14-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
  UI-CF-EN1-acw-1-14-0-UI 3', mRNA sequence.
ACCESSION
  BM975958
VERSION
  BM975958.1 GI:19593549
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 729)
REFERENCE
  AUTHORS
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  JOURNAL
  Genome Res. 6 (9), 791-806 (1996)
  MEDLINE
  97044477
  COMMENT
  Contact: McCray, PB
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
```

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
The following repetitive elements were found in this cDNA  
sequence: 1-35, >AT-rich#low\_complexity  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
Location/Qualifiers  
1..729  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-acw-1-14-0-UI"  
/clone\_lib="UI-CF-EN1"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a  
modified polylinker; Site:1; EcoR I; Site:2; Not I;  
UI-CF-EN1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Cystic Fibrosis  
Epithelial Cells. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoR I adaptor, digested with Not  
I, and cloned directionally into pT7T3-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is CTGCTCAGGT.  
TAG.LIB=UI-CF-EN1  
TAG.TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGGT"

BASE COUNT 171 a 183 c 195 g 179 t 1 others  
ORIGIN  
Query Match 43.1%; Score 610; DB 14; Length 729;  
Best Local Similarity 99.8%; Pred.No. 0;  
Matches 660; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 731 CCCTGAACCTCTCTGCGCATTTGGCTCAAGCTGGAGCAGAGATACCTATCCTTCAGC 790  
|||||  
Db 675 CCCTGAACCTCTCTGCGCATTTGGCTCAAGCTGGAGCAGAGATACCTATCCTTCAGC 616  
|||||  
QY 791 TTTTGGATCCATTACAGGCCACAGCCACTGGGAGGTGTGCTGGGGCCCGTGCTATCGA 850  
|||||  
Db 615 TTNTGGATCCATTACAGGCCACAGCCACTGGGAGGTGTGCTGGGGCCCGTGCTATCGA 556  
|||||  
QY 851 AACCCAGTGTATGTAGTGGCAGCAGCAGTGTGGAGCCACCATTGAGAGAGAGCAG 910  
|||||  
Db 555 AACCCAGTGTATGTAGTGGCAGCAGCAGTGTGGAGCCACCATTGAGAGAGAGCAG 496  
|||||  
QY 911 TTATGCCACACAGCAGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGTCTCTGAGGG 970  
|||||  
Db 495 TTATGCCACACAGCAGTGGTAGACCCCTGGGGAACAGTGGTGGCCCGTCTCTGAGGG 436  
|||||  
QY 971 GCCAGGCTCTGCTTGGCCGGAATAGACTCAACTATCTGCGACAGATTGCGCCGACACT 1030  
|||||  
Db 435 GCCAGGCTCTGCTTGGCCGGAATAGACTCAACTATCTGCGACAGATTGCGCCGACACT 376  
|||||  
QY 1031 GCCTGTGTCCAGCAGCCAGCGCTGACCTCTATGGCAATCTGGGTCAACCACTCTCTTA 1090  
|||||  
Db 375 GCCTGTGTCCAGCAGCCAGCGCTGACCTCTATGGCAATCTGGGTCAACCACTCTCTTA 316  
|||||  
QY 1091 AGACTTGAATCTCTGTGAGTTTAGACCTGCCCTCCACCCACCCCTGCCACTATGAGCT 1150  
|||||  
Db 315 AGACTTGAATCTCTGTGAGTTTAGACCTGCCCTCCACCCACCCCTGCCACTATGAGCT 256  
|||||  
QY 1151 AGTGTCTATGTGACTTGGAGGAGGATCAGGACAGCTGCCCTCAGTTGGAGAACCTTGG 1210  
|||||

Db 255 AGTGCTCATGTGACTGGAGGAGGATCCAGGCACAGCTCCCTCACTGGAGAACCTTG 196

QY 1211 ACTCTCTTGATGAACACAGATGGCTGTGGGAAAGAACTTTACACCTGAGCTTCACC 1270

Db 195 ACTCTCTTGATGAACACAGATGGCTGTGGGAAAGAACTTTACACCTGAGCTTCACC 136

QY 1271 TGAGGTGAGACTCAGTTTCAGAAAGGTGGAATTTATATAGTCATGTTTATTCATGG 1330

Db 135 TGAGGTGAGACTCAGTTTCAGAAAGGTGGAATTTATATAGTCATGTTTATTCATGG 76

QY 1331 AAAGTGAAGTTCGCTGAGGGCTGAGCAGCAGCTGGCATTGAAAATATATATATATATA 1390

Db 75 AAAGTGAAGTTCGCTGAGGGCTGAGCAGCAGCTGGCATTGAAAATATATATATATATA 16

QY 1391 G 1391

Db 15 G 15

RESULT 11

BM690873

LOCUS

DEFINITION 613 bp mRNA linear EST 28-FEB-2002

UI-E-C11-aaz-a-06-0-UI.r1 UI-E-C11 Homo sapiens cDNA clone

UI-E-C11-aaz-a-06-0-UI 5', mRNA sequence.

ACCESSION BM690873

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL

MEDLINE

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..613

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="UI-E-C11-aaz-a-06-0-UI"

/clone\_lib="UI-E-C11"

/tissue\_type="rPE and Choroid"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-C11 is a normalized cDNA library containing the following tissue(s): rPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA.

This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 128 a 184 c 164 g 137 t

ORIGIN

Query Match 42.8%; Score 606; DB 14; Length 613;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 645 TCTACCACCTCGGGCCAGCTCTTGAGTCACCTGTGACACACCCAGCAGCAGGCAAGATTGGT 704

Db 8 TCTACATATCCCTGGGCCCCAGTCTTGAGTCACCTGTGACACACCCAGCAGCAGGCAAGATTGGT 67

QY 705 CTAGCTGTCTGTATGACATGGGGTCCCTGAACTCTCTCTGTCATTGCTCAAGCTGGA 764

Db 68 CTAGCTGTCTGTATGACATGGGGTCCCTGAACTCTCTCTGTCATTGCTCAAGCTGGA 127

QY 765 GCAGAGATACCTTACCTATCTCTTTCAGCTTTTGGATCCATTACAGGCCAGCCCACTGGGAG 824

Db 128 GCAGAGATACCTTACCTATCTCTTTCAGCTTTTGGATCCATTACAGGCCAGCCCACTGGGAG 187

QY 825 GTGTTGCTGGGGCCCGTCTATCGAAACCCAGTCTATGTCAGCAGCAGCAGCAGCAGTGT 884

Db 188 GTGTTGCTGGGGCCCGTCTATCGAAACCCAGTCTATGTCAGCAGCAGCAGCAGTGT 247

QY 885 GGACGCCACCATGAGAAAGAGAGCAAGTTATGGCCACAGCATGTTGTCAGACCCCTGGGGA 944

Db 248 GGACGCCACCATGAGAAAGAGAGCAAGTTATGGCCACAGCATGTTGTCAGACCCCTGGGGA 307

QY 945 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCTCTGCTTGCCTTGCCTGAGTACACCTCAAC 1004

Db 308 ACAGTGGTGGCCCGCTGCTCTGAGGGGCCAGGCTCTGCTTGCCTTGCCTGAGTACACCTCAAC 367

QY 1005 TATCTCGCAGAGTGGCGCCGACACCTGCTGTTTCCAGACCCGAGGCTGACCTCTAT 1064

Db 368 TATCTCGCAGAGTGGCGCCGACACCTGCTGTTTCCAGACCCGAGGCTGACCTCTAT 427

QY 1065 GGCATCTGGGTACACCACTGTCTTAAGACTTGACTTCTGTAGTTTACACCTGCCCCCTC 1124

Db 428 GGCATCTGGGTACACCACTGTCTTAAGACTTGACTTCTGTAGTTTACACCTGCCCCCTC 487

QY 1125 CCACCCCACTCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGGCA 1184

Db 488 CCACCCCACTCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGGCA 547

QY 1185 CAGCTCCCTCACTTGGAGAACCTTGACTCTCTTTGATGGAACACAGATGGGCTGCTTGGG 1244

Db 548 CAGCTCCCTCACTTGGAGAACCTTGACTCTCTTTGATGGAACACAGATGGGCTGCTTGGG 607

QY 1245 AAAGAA 1250

Db 608 AAAGAA 613

RESULT 12

BI769604

LOCUS

DEFINITION

846 bp mRNA linear EST 25-SEP-2001

603054985F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5204469 5', mRNA sequence.

ACCESSION BI769604

VERSION BI769604.1

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Qy	710	TGTC 713      661 TGTC 664
RESULT 13		
LOCUS	BM979227/c	
DEFINITION	750 bp mRNA linear EST 21-MAR-2002 UI-CF-DU1-adl-n-04-0-UI.s1 UT-CF-DU1 Homo sapiens cDNA clone UI-CF-DU1-adl-n-04-0-UI.3', mRNA sequence.	
ACCESSION	BM979227	
VERSION	BM979227.1	GI:19599461
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 750)	
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery	
JOURNAL	Genome Res. 6 (9), 791-806 (1996)	
MEDLINE	9704477	
COMMENT	Contact: McCray, PB McCray Lab University of Iowa 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA Tel. 319 356 4866	

```

Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-41, >AT-rich#Low_complexity
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
    source
        1..750
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="UI-CF-DU1-adl-n-04-0-UI"
            /clone_lib="UI-CF-DU1"
            /tissue_type="Primary Lung Epithelial Cells"
            /dev_stage="Adult"
            /lab_host="DH10B (Life Technologies) ('I1 phage resistant')
            /note="Organ: Lung; Vector: pF7T3-Pac (Pharmacia) with a
            modified polylinker; Site.1: EcoR I; Site.2: Not I;
            UI-CF-DU1 is a normalized cDNA library containing the
            following tissue(s): Primary Lung Epithelial Cells
            The library was constructed according to Bonaldo, Lennon and
            Soares, Genome Research, 6:791-806, 1996. First strand
            cDNA synthesis was primed with an oligo-dT primer
            containing a Not I site. Double stranded cDNA was ligated
            to an EcoR I adaptor, digested with Not I, and cloned
            directionally into pF7T3-Pac vector. The oligonucleotide
            used to prime the synthesis of first-strand cDNA contains
            a library tag sequence that is located between the Not I
            site and the (dT)18 tail. The sequence tag for this
            library is GGCTGTAGGC.
            TAG_LIB=UI-CF-DU1
            TAG_TISSUE=Lung Epithelial Cells
            TAG_SBO=GGCTGTAGGC"
        185 a 171 c 193 g 200 t 1 others

BASE COUNT
ORIGIN

Query Match 41.6%; Score 589; DB 14; Length 750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

	CATGCTGGGCCAGTCTTGAGTCACCTGTGCAGCACACCAGCGCAAGATTGGTTAGC	709
QY	650	
Dδ	601	
	CATGCTGGGCCAAGTCTTGAGTCACCTGTGCAGCACACCAGCGCAAGATTGGTTAGC	660

```
QY 823 AGGTGTTGCTGGGGCCCTGCTGCTATCGAAACCCAGTGCCTATAGTAGGCGAGCACAGT 882
Db 589 AGGTGTTGCTGGGGCCCTGCTGCTATCGAAACCCAGTGCCTATAGTAGGCGAGCACAGT 530
QY 883 GTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGTGGTAGACCCCTGGG 942
Db 529 GTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGTGGTAGACCCCTGGG 470
QY 943 GAACAGTGTGGCCCGCTGCTCTGAGGGGCCAGGCGCTCTGCGCTTGCGCCGAATAGACCTCA 1002
Db 469 GAACAGTGTGGCCCGCTGCTCTGAGGGGCCAGGCGCTCTGCGCTTGCGCCGAATAGACCTCA 410
QY 1003 ACTATCTGGACAGATGGCCGACACCTGCCCTGTGTTCCAGCACCGCAGGCGTGAACCTT 1062
Db 409 ACTATCTGGACAGATGGCCGACACCTGCCCTGTGTTCCAGCACCGCAGGCGTGAACCTT 350
QY 1063 ATGGCAATCTGGGTACACCACTGCTTAAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 1122
Db 349 ATGGCAATCTGGGTACACCACTGCTTAAAGACTTGACTTCTGTGAGTTTAGACCTGCCCC 290
QY 1123 TCCACCCCCACCCCTGCCACTATGACTAGTGTCTATGTGACTTGGAGCAGGATCCAGG 1182
Db 289 TCCACCCCCACCCCTGCCACTATGACTAGTGTCTATGTGACTTGGAGCAGGATCCAGG 230
QY 1183 CACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 1242
Db 229 CACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGGCTGCTTG 170
QY 1243 GGAAGAAACTTTCACCTGAGCTTACCTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1302
Db 169 GGAAGAAACTTTCACCTGAGCTTACCTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 110
QY 1303 TTTTATATAGTCATTTTATTTATTCATGGAACCTGAACTTCTGCTGAGGCTGAGCAGCAC 1362
Db 109 TTTTATATAGTCATTTTATTTATTCATGGAACCTGAACTTCTGCTGAGGCTGAGCAGCAC 50
QY 1363 TGGCATTGAAAAATATAATAATCATAAAGTCAAAAAAATCAAAAAAATCAAAAAA 1411
Db 49 TGGCATTGAAAAATATAATAATCATAAAGTCAAAAAAATCAAAAAAATCAAAAAA 1

RESULT 14
AL520767/c
LOCUS
DEFINITION
AL520767 LTI_NFL004_NBC2 Homo sapiens cDNA clone EST 13-FEB-2001
prime, mRNA sequence.
ACCESSION
AL520767
VERSION
AL520767.1 GI:12784260
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 92)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB002YD05"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="PH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
```

```
BASE COUNT 229 a 267 c 268 g 222 t 6 others
ORIGIN
Query Match 41.3%; Score 585; DB 9; Length 992;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 805; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 443 ATACACCCAGCTTGCAGGGAATGTGGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCG 502
Db 929 ATACACCCAGCTTGCAGGGAATGTGGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCG 870
QY 503 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTTCTACGTCGCTCTGAACAGCAA 562
Db 869 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTTCTACGTCGCTCTGAACAGCAA 810
QY 563 AGGGCAGTAGTGGCCACTTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 622
Db 809 AGGGCAGTAGTGGCCACTTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 750
QY 623 GGGGCTATGTCTGAAAGCAACTCTACCATGCTGGGCCAGCTCTTGACTCACTGTCAG 682
Db 749 GGGGCTATGTCTGAAAGCAACTCTACCATGCTGGGCCAGCTCTTGACTCACTGTCAG 690
QY 683 CACACAGCAGCAAGATTGGTCTAGCTGTCTATGACATGCGGTTCCTTGAACCTCTC 742
Db 689 CACACAGCAGCAAGATTGGTCTAGCTGTCTATGACATGCGGTTCCTTGAACCTCTC 630
QY 743 TCTGGCATTGGCTCAAGCTGGAGCAGAGATCTTACCTATCCTTCAGCTTTTGGATCCAT 802
Db 629 TCTGGCATTGGCTCAAGCTGGAGCAGAGATCTTACCTATCCTTCAGCTTTTGGATCCAT 570
QY 803 TACAGGCCAGCCCACTGGGAGGTGTGCTGGGCCGCTGTATCGAAACCCAGTGCTA 862
Db 569 TACAGGCCAGCCCACTGGGAGGTGTGCTGGGCCGCTGTATCGAAACCCAGTGCTA 510
QY 863 TGTAGTGGC - AGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 921
Db 509 TGTAGTGGCAGCAGCACAGTGTGGAGGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 450
QY 922 GCATGGTGTACACCCCTGGGGAACAGTGGTGGCCGCTGCTCTGAGGGGCCAGGCCTCT 981
Db 449 GCATGGTGTACACCCCTGGGGAACAGTGGTGGCCGCTGCTCTGAGGGGCCAGGCCTCT 390
QY 982 GCCTTGGCCGAATAGACCTCAACTATCTCGACAGTGTGGCCGACACCTGCTGTGTTCC 1041
Db 389 GCCTTGGCCGAATAGACCTCAACTATCTCGACAGTGTGGCCGACACCTGCTGTGTTCC 330
QY 1042 AGCACCGCAGGCTGACCTCTATGCGCAATCTGGGTCAACCCACTGTCTTAAGACTTTGACTT 1101
Db 329 AGCACCGCAGGCTGACCTCTATGCGCAATCTGGGTCAACCCACTGTCTTAAGACTTTGACTT 270
QY 1102 CTGTGAGTTTACACCTGCCCCCTCCACCCGCCCTGACCTATGAGCTAGTGCATGT 1161
Db 269 CTGTGAGTTTACACCTGCCCCCTCCACCCGCCCTGACCTATGAGCTAGTGCATGT 210
QY 1162 GACTTGGAGGAGGATCCAGGACAGCTCCCTCACTTTGGAACCTTGACTCTCTTGAT 1221
Db 209 GACTTGGAGGAGGATCCAGGACAGCTCCCTCACTTTGGAAGAA 1249
QY 1222 GGAACACAGATGGGCTGCTTGGGAAGA 1249
Db 149 GGAACACAGATGGGCTGCTTGGGAAGA 122
RESULT 15
AT1797259/c
```

cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetechn.com URL : http://fulllength.invitrogen.com"

LOCUS AI797259 586 bp mRNA linear EST 18-DEC-1999  
DEFINITION we85e02.x1 Soares NFL\_T.GBC.S1 Homo sapiens cDNA clone  
IMAGE:2347994 3' similar to TR:076091 076091 NITRILASE HOMOLOG 1.  
; mRNA sequence.  
ACCESSION AI797259  
VERSION AI797259.1 GI:5362731  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 586)  
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1500 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 468.  
Location/Qualifiers  
1..586  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2347994"  
/clone\_lib="Soares\_NFL\_T.GBC.S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI-CGAP-GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Ronaldo."

FEATURES  
source

BASE COUNT 137 a 155 c 158 g 136 t  
ORIGIN  
Query Match 41.2%; Score 584; DB 9; Length 586;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 810 CCAGCCCACTGGGAGGTGTGCTGCGGGCCCGTGTATCGAAACCCAGTGCCTATGTAGTG 869  
Db 586 CCAGCCCACTGGGAGGTGTGCTGCGGGCCCGTGTATCGAAACCCAGTGCCTATGTAGTG 527  
QY 870 GCAGCAGCACAGTGTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGGTG 929  
Db 526 GCAGCAGCACAGTGTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGGTG 467  
QY 930 GTAGACCCCTGGGACACAGTGTGGCCCGCTGCTCTGAGGGCCAGGCCCTCGCCTTGCC 989  
Db 466 GTAGACCCCTGGGAAACAGTGTGGCCCGCTGCTCTGAGGGCCAGGCCCTCGCCTTGCC 407  
QY 990 CGAATAGACCTCAACTATCTGGGACAGTTGGCGCCGACACCTGCCTGTGTCCAGCACCGC 1049  
Db 406 CGAATAGACCTCAACTATCTGGGACAGTTGGCGCCGACACCTGCCTGTGTCCAGCACCGC 347  
QY 1050 AGGCCTGACCTCTATGGCAATCTGGGTACACCACCTGTCTTAAGACTTGACTTCTGTGAGT 1109  
Db 346 AGGCCTGACCTCTATGGCAATCTGGGTACACCACCTGTCTTAAGACTTGACTTCTGTGAGT 287  
QY 1110 TTAGACCTGCCCTCCACCCACCCTGCCACTATGAGCTAGTGCCTCATGTGACTTGA 1169  
Db 286 TTAGACCTGCCCTCCACCCACCCTGCCACTATGAGCTAGTGCCTCATGTGACTTGA 227

QY 1170 GGCAGGATCCAGGCACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGTATGGAACACA 1229  
Db 226 GGCAGGATCCAGGCACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGTATGGAACACA 167  
QY 1230 GATGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTGAGACTGCAGTTT 1289  
Db 166 GATGGCTGCTTGGGAAAGAAACTTTCACCTGAGCTTCACCTGAGGTGAGACTGCAGTTT 107  
QY 1290 CAGAAAGGTGGAAATTTATATAGTCATTGTTTATTTTCATGGAAACTGAAGTTCTGCTGAG 1349  
Db 106 CAGAAAGGTGGAAATTTATATAGTCATTGTTTATTTTCATGGAAACTGAAGTTCTGCTGAG 47  
QY 1350 GGCTGAGCAGCACTGGCAATTTGAAAAATATAATATCAATAAGTC 1393  
Db 46 GGCTGAGCAGCACTGGCAATTTGAAAAATATAATATCAATAAGTC 3

Search completed: December 14, 2002, 18:30:06  
Job time : 2043 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:37:34 ; Search time 2740 Seconds  
(without alignments)  
15039.990 Million cell updates/sec

Title: US-09-357-675C-1  
Perfect score: 1416  
Sequence: 1 gcccaactcgctgcgcctct.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pt.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_hugo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1307	92.3	1385	9	AF069987	AF069987 Homo sapi
2	571	40.3	4079	9	AF069984	AF069984 Homo sapi
c 3	571	40.3	167863	2	AL590651	AL590651 Homo sapi
4	571	40.3	200822	9	AL591806	AL591806 Human DNA
c 5	326	23.0	377	6	AX368386	AX368386 Sequence
c 6	173	12.2	441	11	G13436	G13436 human STS W
c 7	87	6.1	2261	6	AX397469	AX397469 Sequence
c 8	87	6.1	2261	9	AK001497	AK001497 Homo sapi
9	47	3.3	847	11	G72919	G72919 MARC 2849-2
10	33	2.3	1338	10	AF069988	AF069988 Mus muscu
11	33	2.3	1365	10	BC021634	BC021634 Mus muscu
12	33	2.3	4481	10	AF069985	AF069985 Mus muscu
13	33	2.3	181583	2	AC105589	AC105589 Rattus no
14	33	2.3	211772	10	AC084821	AC084821 Mus muscu
c 15	33	2.3	215043	10	AC087229	AC087229 Mus muscu
c 16	32	2.3	115667	2	AC125857	AC125857 Rattus no
17	31	2.2	144986	2	AC108571	AC108571 Rattus no
18	30	2.1	1587	9	BC017413	BC017413 Homo sapi
19	30	2.1	2421	10	BC016425	BC016425 Mus muscu
20	30	2.1	3300	6	AR212407	AR212407 Sequence
21	30	2.1	4113	10	AF230376	AF230376 Meriones
c 22	30	2.1	44645	9	AC108058	AC108058 Homo sapi
23	30	2.1	60773	2	AC069576	AC069576 Homo sapi
c 24	30	2.1	252059	2	AC102081	AC102081 Mus muscu
25	29	2.0	345	10	ENRIPRL38	X57007 Rat mRNA fo
26	29	2.0	549	9	AF112344	AF112344 Homo sapi
c 27	29	2.0	564	6	AX407609	AX407609 Sequence
28	29	2.0	876	5	AF175983	AF175983 Rana sylv
29	29	2.0	1049	8	AF071889	AF071889 Prunus ar
30	29	2.0	1075	9	BC014385	BC014385 Homo sapi
31	29	2.0	1134	10	AF058799	AF058799 Mus muscu
32	29	2.0	1193	6	A29423	A29423 putative bo
33	29	2.0	1193	6	AR022483	AR022483 Sequence
34	29	2.0	1193	6	AR068961	AR068961 Sequence
35	29	2.0	1193	6	AR103330	AR103330 Sequence
36	29	2.0	1193	6	AR134988	AR134988 Sequence
37	29	2.0	1193	6	AR141817	AR141817 Sequence
38	29	2.0	1193	6	AR143334	AR143334 Sequence
39	29	2.0	1193	6	AR151943	AR151943 Sequence
40	29	2.0	1193	6	I36423	I36423 Sequence 13
41	29	2.0	1193	6	I88123	I88123 Sequence 13
42	29	2.0	1211	5	AY065841	AY065841 Danio rer
43	29	2.0	1267	10	BC004064	BC004064 Mus muscu
44	29	2.0	1488	9	HSU79258	U79258 Human clone
45	29	2.0	1667	3	AY047568	AY047568 Drosophil

ALIGNMENTS

RESULT 1  
AF069987  
LOCUS Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.  
DEFINITION Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.  
ACCESSION AF069987  
VERSION AF069987.1 GI:3228665  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1385)  
AUTHORS Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,  
Tillib, S., Draganes, A., Wermuth, P., Rothman, J. H., Huebner, K.,  
Buchberg, A. M., Mazo, A., Brenner, C. and Croce, C. M.





JOURNAL Drosophila melanogaster and Caenorhabditis elegans  
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
REFERENCE 98337986  
PUBMED 9671749  
AUTHORS 2 (bases 1 to 4079)  
Pekarsky, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,  
Tillib, S., Draganes, A., Werbuth, P., Rothman, J., Huebner, K.,  
Buchberg, A.M., Mazon, A., Brenner, C. and Croce, C.M.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA  
FEATURES Location/Qualifiers  
source  
1..4079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21-q22"  
635..3569  
/gene="NIT1"  
-join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3569)  
/gene="NIT1"  
/product="nitric oxide synthase homolog 1"  
/note="alternatively spliced"  
635..680  
/gene="NIT1"  
/note="alternatively spliced"  
/number=1  
-join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3267)  
/gene="NIT1"  
/note="NIT1; alternatively spliced"  
/codon\_start=1  
/product="nitric oxide synthase homolog 1"  
/protein\_id="AAC39901.1"  
/db\_xref="GI:3242378"  
/translation="MLGFTIPPPHRLSLCPLRIPOLSVLCAPRPRMAISSSS  
EPLVAVQVSTPDKQNFKTCALVREARLGLACFLPEAFDFIARDPAETLHLS  
EPLGKLEETQLAREGLMLSLGPFHREGDWEQTKIYNCHVLLNKGAVVATYR  
KTHLGDIPGPGCESNTPGPSLESPVTPAGKIGLAVCYDMRPPELSALAOA  
GAEILTYPSGSIQPAHWEVLLRARETQCYVVAACQGRHHEKRSYGHSMVD  
PAGTVACSEPGICLARDLNLRLRHLRFQHRPDDIYGLHGLPLS"  
681..711  
/gene="NIT1"  
/note="1A; alternatively spliced"  
712..741  
/gene="NIT1"  
/note="1B; alternatively spliced"  
940..1053  
/gene="NIT1"  
/note="1C; alternatively spliced"  
1288..1383  
/gene="NIT1"  
/number=2  
1636..1890  
/gene="NIT1"  
/number=3  
2011..2114  
/gene="NIT1"  
/number=4  
2321..2454  
/gene="NIT1"  
/number=5  
2609..2734  
/gene="NIT1"  
/number=6  
3001..3569  
/gene="NIT1"  
/number=7  
940 a 1082 c 1043 g 1012 t 2 others

BASE COUNT  
ORIGIN

Query Match

40.3%; Score 571; DB 9; Length 4079;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 823 AGTGTTGTCGGGCGCGTGTATCGAAGCCAGTGTCTAGTGGCAGCAGCAGT 882  
|||||  
Db 2999 AGTGTTGTCGGGCGCGTGTATCGAAGCCAGTGTCTAGTGGCAGCAGCAGT 3058  
QY 883 GTGGAGCCACCATTGAGAGAGAGCAAGTTATGGCCACAGCAGTGTGTAGACCCCTGGG 942  
|||||  
Db 3059 GTGGAGCCACCATTGAGAGAGAGCAAGTTATGGCCACAGCAGTGTGTAGACCCCTGGG 3118  
QY 943 GAACAGTGTGGCCCGCTGCTGTGAGGGCCAGGCTCTGCTTGTCCAGCAGCAGCAGCCTGACCTCT 1002  
|||||  
Db 3119 GAACAGTGTGGCCCGCTGCTGTGAGGGCCAGGCTCTGCTTGTCCAGCAGCAGCAGCCTGACCTCT 3178  
QY 1003 ACTATCTCGACAGTTGCGCCGACACCTGCTGTGTCCAGCAGCAGCAGCCTGACCTCT 1062  
|||||  
Db 3179 ACTATCTCGACAGTTGCGCCGACACCTGCTGTGTCCAGCAGCAGCAGCCTGACCTCT 3238  
QY 1063 ATGGCAATCTGGTCAACCCACTGCTTTAAGACTTGTCTGTGAGTTTAGACTTGCCTC 1122  
Db 3239 ATGGCAATCTGGTCAACCCACTGCTTTAAGACTTGTCTGTGAGTTTAGACTTGCCTC 3298  
QY 1123 TCCACCCCGCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTTGGAGCAGGATCCAGG 1182  
Db 3299 TCCACCCCGCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTTGGAGCAGGATCCAGG 3358  
QY 1183 CACAGTCTCCCTCAGTGGAGAACCTTGTGCTCTTTGATGGAAACAGATGGGCTGCTTG 1242  
Db 3359 CACAGTCTCCCTCAGTGGAGAACCTTGTGCTCTTTGATGGAAACAGATGGGCTGCTTG 3418  
QY 1243 GGAAGAAACTTTACCTGAGCTTACCTGAGGTGACAGTTCAGAAAGGTGGAA 1302  
Db 3419 GGAAGAAACTTTACCTGAGCTTACCTGAGGTGACAGTTCAGAAAGGTGGAA 3478  
QY 1303 TTTTATATAGTCAATGTTTATTTATGGAACCTGAAGTTCCTGCTGAGGCGGTGAGCAGCAC 1362  
Db 3479 TTTTATATAGTCAATGTTTATTTATGGAACCTGAAGTTCCTGCTGAGGCGGTGAGCAGCAC 3538  
QY 1363 TGGCAATGAAATATAATAATCATAAAGTC 1393  
Db 3539 TGGCAATGAAATATAATAATCATAAAGTC 3569

RESULT 3

AL590651/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL590651 Homo sapiens chromosome 1 clone RP11-137A12, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 9 unordered pieces.  
AL590651 AL590651.4 GI:13990236  
HTG: HTGS\_PHASE1; HTGS\_CANCELLED.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 167863)  
Harrison, E.  
Direct Submission  
Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On May 7, 2001 this sequence version replaced gi:13751001.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BAI37A12  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 164729 bases at least Q40  
Consensus quality: 165486 bases at least Q30  
Consensus quality: 166169 bases at least Q20  
Insert size: 167063; sum-of-contigs  
Insert size: 171878; 3.4% error; agarose-fp  
Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality  
coverage: 8.11x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 9374 9473: contig of 9373 bp in length  
\* 9474 29891: contig of 20418 bp in length  
\* 29892 29991: gap of 100 bp  
\* 29992 36239: contig of 6248 bp in length  
\* 36240 36339: gap of 100 bp  
\* 36340 103784: contig of 67445 bp in length  
\* 103785 103884: gap of 100 bp  
\* 103885 114878: contig of 10994 bp in length  
\* 114879 114978: gap of 100 bp  
\* 114979 126347: contig of 11369 bp in length  
\* 126348 126447: gap of 100 bp  
\* 126448 137764: contig of 11317 bp in length  
\* 137765 137864: gap of 100 bp  
\* 137865 163160: contig of 25296 bp in length  
\* 163161 163260: gap of 100 bp  
\* 163261 167863: contig of 4603 bp in length.

FEATURES

source

1..167863  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-137A12"  
/clone\_lib="RPC1-11.1"

misc\_feature

1..9373  
/note="assembly\_fragment:02360  
fragment\_chain:1  
clone\_end:T7  
vector\_side:left"

misc\_feature

9474..29891  
/note="assembly\_fragment:01854  
fragment\_chain:1"

misc\_feature

29992..36239  
/note="assembly\_fragment:00805  
fragment\_chain:1"

misc\_feature

36340..103784  
/note="assembly\_fragment:01467  
fragment\_chain:1"

misc\_feature

103885..114878  
/note="assembly\_fragment:00991  
fragment\_chain:2"

misc\_feature

114979..126347  
/note="assembly\_fragment:01503  
fragment\_chain:2"

misc\_feature

126448..137764  
/note="assembly\_fragment:01297  
fragment\_chain:2"

misc\_feature

137865..163160  
/note="assembly\_fragment:00117  
fragment\_chain:2"

misc\_feature

163261..167863  
/note="assembly\_fragment:01267  
fragment\_chain:2  
clone\_end:SP6  
vector\_side:right"

BASE COUNT 43927 a 39935 c 40002 g 43197 t 802 others

ORIGIN

Query Match 40.3%; Score 571; DB 2; Length 167863;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 823 AGGTGTTGTCGGGGCCCGTGTATCGAAACCCAGTGTATGTAGTGGCAGCAGCACACT 882  
|||||  
Db 149682 AGGTGTTGTCGGGGCCCGTGTATCGAAACCCAGTGTATGTAGTGGCAGCAGCACACT 149623  
QY 883 GTGGAGCCACCATTGAGAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 942  
|||||  
Db 149622 GTGGAGCCACCATTGAGAGAGAGCAAGTTATGGCCACAGCATGGTGGTAGACCCCTGGG 149563  
QY 943 GAACAGTGGTGGCCCGCTCTGTAGGGCCAGGCTCTGCCTTGGCCCAATAGACCTCA 1002  
|||||  
Db 149562 GAACAGTGGTGGCCCGCTCTGTAGGGCCAGGCTCTGCCTTGGCCCAATAGACCTCA 149503  
QY 1003 ACTATCTGCGACAGTTGGCCCGACACCTGCTCTGTGTTCCAGCACCGCAGCCTGACCTCT 1062  
|||||  
Db 149502 ACTATCTGCGACAGTTGGCCCGACACCTGCTCTGTGTTCCAGCACCGCAGCCTGACCTCT 149443  
QY 1063 ATGGCAATCTGGGTCAACCCACTGTCTTAAGACTTGACTTCTGTAGTTTAGACCTGCCCC 1122  
|||||  
Db 149442 ATGGCAATCTGGGTCAACCCACTGTCTTAAGACTTGACTTCTGTAGTTTAGACCTGCCCC 149383  
QY 1123 TCCACCCCGCCCGCTGCTATGAGCTAGTCTCATGTGACTTGGAGCAGGATCCAGG 1182  
|||||  
Db 149382 TCCACCCCGCCCGCTGCTATGAGCTAGTCTCATGTGACTTGGAGCAGGATCCAGG 149323  
QY 1183 CACAGTCCCTCACTTGGAGAACCTTGACTCTCTTTGATGGAACACAGATGGCTGCTTG 1242  
|||||  
Db 149322 CACAGTCCCTCACTTGGAGAACCTTGACTCTCTTTGATGGAACACAGATGGCTGCTTG 149263  
QY 1243 GGAAGAAACTTTCACCTGAGCTTCCACCTGAGCTGAGCTGAGCTTTCAGAAAGTGGAA 1302  
|||||  
Db 149262 GGAAGAAACTTTCACCTGAGCTTCCACCTGAGCTGAGCTGAGCTTTCAGAAAGTGGAA 149203  
QY 1303 TTTTATATAGTCATGTTTATTTTATGGAACCTGAAGTTCTCTGAGGCTCAGCAGCAC 1362  
|||||  
Db 149202 TTTTATATAGTCATGTTTATTTTATGGAACCTGAAGTTCTCTGAGGCTCAGCAGCAC 149143  
QY 1363 TGGCATTTGAAAAATATAATATATAAAGTC 1393  
|||||  
Db 149142 TGGCATTTGAAAAATATAATATAAAGTC 149112

RESULT 4

AL591806

LOCUS

Human DNA sequence from clone RP11-544M22 on chromosome 1, complete  
sequence.

ACCESSION

AL591806

VERSION

AL591806.16

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk  
On Feb 1, 2002 this sequence version replaced gi:17902927.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone, and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> Rpl1-544M22 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

THIS sequence is the entire insert of clone RPl1-544M22 The true left end of clone RPl1-137A12 is at 156538 in this sequence. The true right end of clone RPl1-381D2 is at 145015 in this sequence.

FEATURES  
source  
1..200822  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RPl1-544M22"  
/clone\_lib="RPl1-11.2"  
10395..10397  
/note="Sequence from overlapping clone RPl1-381D2 (AL162592). Assembly confirmed by restriction digest."  
44098..44107  
/note="Sequence from overlapping clone RPl1-381D2 (AL162592). Assembly confirmed by restriction digest."  
59585..59604  
/note="Sequence from overlapping clone RPl1-381D2 (AL162592). Assembly confirmed by restriction digest."  
90797..90883  
/note="Single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."  
90861..90883  
/note="Sequence from uni-directional dGTP big dye terminator reads only."  
BASE COUNT 55215 a 49052 c 46183 g 50372 t  
ORIGIN  
Query Match 40.3%; Score 571; DB 9; Length 200822;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 571; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 823 AGGTGTTGCTGGGGCCCGTCTATCGAACCCAGTGTATGTAGTGGCAGGACAGT 882  
Db 174499 AGGTGTTGCTGGGGCCCGTCTATCGAACCCAGTGTATGTAGTGGCAGGACAGT 174558  
QY 883 GTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGAGTGTGGTAGACCCCTGGG 942  
Db 174559 GTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGAGTGTGGTAGACCCCTGGG 174618  
QY 943 GAACAGTGTGGCCCGCTGCTCTGAGGGCCAGGCCCTCTGCCTTGGCCCAATAGACCTCA 1002  
Db 174619 GAACAGTGTGGCCCGCTGCTCTGAGGGCCAGGCCCTCTGCCTTGGCCCAATAGACCTCA 174678  
QY 1003 ACTATCTGGACAGTTGGCCGACACCTGCTGTGTTCAGCACCCAGCCGCTGACCTCT 1062  
Db 174679 ACTATCTGGACAGTTGGCCGACACCTGCTGTGTTCAGCACCCAGCCGCTGACCTCT 174738  
QY 1063 ATGGCAATCTGGGTCAACCCACTGCTTTAAGACTTTGACTTCTGTGAGTTAGACCTGCCCC 1122  
Db 174739 ATGGCAATCTGGGTCAACCCACTGCTTTAAGACTTTGACTTCTGTGAGTTAGACCTGCCCC 174798  
QY 1123 TCCACCCCCACCCCTGCCACTATGAGCTAGTGTCTATGTGACTTTGGAGGCGAGGATCCAGG 1182

Db 174799 TCCACCCCCACCCCTGCCACTATGAGCTAGTGTCTATGTGACTTTGGAGCGAGGATCCAGG 174858  
QY 1183 CACAGCTCCCTCCTCAGTGGAGAACCTTGACTCTCTTTGATGGACACAGATGGCTGCTTG 1242  
Db 174859 CACAGCTCCCTCCTCAGTGGAGAACCTTGACTCTCTTTGATGGACACAGATGGCTGCTTG 174918  
QY 1243 GGAAGAAACTTTACCTCAGCTGAGCTTACCTTGAGGTGAGCTGAGCTTTCAGAAAGTGGAA 1302  
Db 174919 GGAAGAAACTTTACCTCAGCTGAGCTTACCTTGAGGTGAGCTGAGCTTTCAGAAAGTGGAA 174978  
QY 1303 TTTTATATAGTCATGTTTATTTATGAGAACTGAAGTTCTGCTGAGGCTGAGCAGCAC 1362  
Db 174979 TTTTATATAGTCATGTTTATTTATTTATGAGAACTGAAGTTCTGCTGAGGCTGAGCAGCAC 175038  
QY 1363 TGGCATTTGAAAATATAATAATCATAAAGTC 1393  
Db 175039 TGGCATTTGAAAATATAATAATCATAAAGTC 175069  
RESULT 5  
LOCUS AX368386/c 377 bp DNA linear PAT 16-FEB-2002  
DEFINITION Sequence 1096 from Patent WO0204514.  
ACCESSION AX368386  
VERSION AX368386.1 GI:18856459  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Wang, T., Watanabe, Y., Henderson, R.A., Johnson, J.C., Retter, M.W., Marnerakis, M., Carter, D., Fanger, G.R., Vedvick, T.S., Bangur, C.S., McInabb, A., Fanger, N., Switzer, A., McNeill, P.D. and Clapper, J.D.  
TITLE Compositions and methods for the therapy and diagnosis of lung cancer  
JOURNAL Patent: WO 0204514-A 1096 17-JAN-2002;  
CORIXA CORPORATION (US)  
FEATURES  
source  
1..377  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 94 a 100 c 112 g 71 t  
ORIGIN  
Query Match 23.0%; Score 326; DB 6; Length 377;  
Best Local Similarity 99.7%; Pred. No. 1.7e-190;  
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 129 CTTCTCTCAGAGATTCCTGTCCTTCTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 188  
Db 377 CTTCTCTCAGAGATTCCTGTCCTTCTGTCCTGGACTCCGGATACCTCAACTCTCAGTA 318  
QY 189 CTTTGTGCTCAGCCAGCCAGGACCATGGCTATCTCTCTTCTCTCCGCAACTGCC 248  
Db 317 CTTTGTGCTCAGCCAGCCAGGACCATGGCTATCTCTCTTCTCTCCGCAACTGCC 258  
QY 249 CTGGTGGCTGTGTGCCAGGTAACTCGACGCCAGACAGCAAGCAACATTTAAACATGT 308  
Db 257 CTGGTGGCTGTGTGCCAGGTAACTCGACGCCAGACAGCAAGCAACATTTAAACATGT 198  
QY 309 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCCTGCTTCTGCTTCTGCTGAGGCA 368  
Db 197 GCTGGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCCTGCTTCTGCTTCTGCTGAGGCA 138  
QY 369 TTTGACTTCATTTCACGGGACCTTCAGAGACGCTACACCTGTCTGAACACACTGGGTGG 428  
Db 137 TTTGACTTCATTTCACGGGACCTTCAGAGACGCTACACCTGTCTGAACACACTGGGTGG 78  
QY 429 AAACCTTTGGAAGAATACACCCAGCTTGCAGGGAATGTGGACTCTGGCTGTCTCTTGGGT 488  
Db 77 AAACCTTTGGAAGAATACACCCAGCTTGCAGGGAATGTGGACTCTGGCTGTCTCTTGGGT 18

QY 489 GGTTCATGAGCGTGG 505  
|||||  
Db 17 GGTTCATGAGCGTGG 1  
|||||

RESULT 6  
G13436/c  
LOCUS human STS WI-12338, sequence tagged site.  
DEFINITION G13436  
ACCESSION G13436  
VERSION G13436.1 GI:1127545  
KEYWORDS STS; STS sequence; primer; sequence tagged site.  
SOURCE Homo sapiens STS derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 441)  
Hudson.T.  
Whitehead Institute/MIT Center for Genome Research; Physically  
Mapped STS  
Unpublished (1995)

CONTACT: Thomas Hudson  
Whitehead Institute/MIT Center for Genome Research  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA  
Tel: 617 252 1900  
Fax: 617 252 1902  
Email: thudson@genome.wi.mit.edu

Primer A: TATATTTTCAATGCCAGTGTG  
Primer B: AACTTTCACGTGAGTTCACC  
STS size: 130  
PCR Profile:  
  Presoak:  
    Denaturation:  
      Annealing: 56 degrees C  
    Polymerization:  
      PCR Cycles: 35  
    Thermal Cycler:  
Protocol:  
  Template: 10 ng  
  Primer: each 5 pM  
  dNTPs: each 4 nM  
  Taq Polymerase: 0.025 units/ul  
  Total Vol: 20 ul

Buffer:  
  MgCl2: 1.5 mM  
  KCl: 50 mM  
  Tris-HCl: 10 mM  
  pH: 9.3

Derived from dbEST (genbank accession T78621).  
LOCATION/Qualifiers  
  1..441  
  /organism="Homo sapiens"  
  /db\_xref="taxon:9606"  
  /map="750.8 cR from top of Chr1 linkage group"  
STS  
  primer\_bind 16..145  
  primer\_bind complement(124..145)  
BASE COUNT 107 a 95 c 130 g 103 t 6 others  
ORIGIN

Query Match 12.2%; Score 173; DB 11; Length 441;  
Best Local Similarity 100.0%; Pred. No. 2e-95;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1178 CCAGGCACAGCTCCCTCAGTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGCT 1237  
|||||

Db 216 CCAGGCACAGCTCCCTCAGTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGCT 157  
QY 1238 GCTTGGGAAAGAACTTTACCTGAGCTTCACTGAGGTCAGACTGCAGCTTTTCAGAAAGG 1297  
|||||  
Db 156 GCTTGGGAAAGAACTTTACCTGAGCTTCACTGAGGTCAGACTGCAGCTTTTCAGAAAGG 97  
|||||

QY 1298 TGGAAATTTTATATAGTCATGTTTATTTTCATGGAACACTGAAGTCTCTGCTGAGG 1350  
|||||  
Db 96 TGGAAATTTTATATAGTCATGTTTATTTTCATGGAACACTGAAGTCTCTGCTGAGG 44  
|||||

RESULT 7  
AX397469/c  
LOCUS AX397469 2261 bp DNA linear PAT 18-MAY-2002  
DEFINITION Sequence 1684 from Patent WO0212328.  
ACCESSION AX397469  
VERSION AX397469.1 GI:21068216  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
King, G.E., Meagher, M.J., Xu, J. and Secrist, H.  
Compositions and methods for the therapy and diagnosis of colon  
cancer  
JOURNAL Patent: WO 0212328-A 1684 14-FEB-2002;  
CORIXA CORPORATION (US)  
FEATURES  
  Location/Qualifiers  
    1..2261  
    /organism="Homo sapiens"  
    /db\_xref="taxon:9606"  
BASE COUNT 543 a 594 c 541 g 583 t  
ORIGIN

Query Match 6.1%; Score 87; DB 6; Length 2261;  
Best Local Similarity 100.0%; Pred. No. 7.1e-42;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 ATATAGTCATGTTTATTTTCATGGAACACTGAAGTCTCTGCTGAGGCTGAGCAGCACTGGC 1366  
|||||  
Db 2261 ATATAGTCATGTTTATTTTCATGGAACACTGAAGTCTCTGCTGAGGCTGAGCAGCACTGGC 2202  
|||||

QY 1367 ATTGAAAATATAATATCAATAAGTC 1393  
|||||  
Db 2201 ATTGAAAATATAATATCAATAAGTC 2175  
|||||

RESULT 8  
AK001497/c  
LOCUS AK001497 2261 bp mRNA linear PRI 01-AUG-2002  
DEFINITION Homo sapiens cDNA FLJ10635 fis, clone NT2RP2005669, highly similar  
to Homo sapiens mRNA for DEDD protein.  
ACCESSION AK001497  
VERSION AK001497.1 GI:7022790  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens teratocarcinoma cell line: NT2 cDNA to mRNA,  
clone\_lib: NT2RP2 clone: NT2RP2005669.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,  
Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K.,  
Tanai, H., Kimata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y.,  
Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K.,  
Masuho, Y. and Kanehori, K.  
NEBO human cDNA sequencing project  
Unpublished  
JOURNAL 2 (bases 1 to 2261)  
AUTHORS Isogai, T. and Otsuki, T.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomicsr1.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing and clone selection;  
Helix Research Institute (supported by Japan Key Technology Center  
etc.) and Department of Virology, Institute of Medical Science,  
University of Tokyo.

FEATURES  
Source  
1. 2261  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP2005669"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP2"  
/note="Cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."

BASE COUNT 543 a 594 c 541 g 583 t  
ORIGIN  
Query Match 6.1%; Score 87; DB 9; Length 2261;  
Best Local Similarity 100.0%; Pred. No. 7.1e-42;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 ATATAGTCATTCTTTATTCATGAAACCTGAAGTCTGCTGAGGCTGAGCAGCACTGGC 1366  
|||||

Db 2261 ATATAGTCATTCTTTATTCATGAAACCTGAAGTCTGCTGAGGCTGAGCAGCACTGGC 2202  
|||||

QY 1367 ATTGAAAATATAATCAATCAAAAGTC 1393  
|||||

Db 2201 ATTGAAAATATAATCAATCAAAAGTC 2175  
|||||

RESULT 9  
G72919  
LOCUS 847 bp DNA linear STS 08-AUG-2001  
DEFINITION MARC 2849-2850:991933517:1 SCF - porcine spleen Sus scrofa STS  
genomic, sequence tagged site.

ACCESSION G72919  
VERSION G72919.1 GI:15146949  
KEYWORDS STS.

SOURCE Sus scrofa  
Sus scrofa  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactylia; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 847)  
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L. and  
Keele,J.W.  
TITLE Single nucleotide polymorphism (SNP) discovery in expressed porcine  
genes  
JOURNAL Unpublished  
COMMENT

Contact: Freking BA  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4278  
Fax: 402 762 4173  
Email: freking@email.marc.usda.gov  
Primer A: GGGTTCAGCTTTGGATCTG  
Primer B: GCTGAGCCTGTGTTTCATCA  
STS size: 700  
PCR Profile:

Hotstart: 95 degrees for 15 minutes  
Denature: 95 degrees for 30 seconds  
Anneal: 58 degrees  
Extension: 68 degrees for 2 minutes  
Cycles: 32 to 45

Protocol:  
Template: 50-200 ng genomic DNA  
Primer: each 20 pmol

dNTPs: each 88 uM  
Taq Polymerase: 0.25 units (Qiagen HotStar)  
Buffer:  
Commercially supplied Qiagen HotStar buffer

The STS is derived from PCR amplicons generated from genomic DNA,  
sequenced from each end using the amplification primers. The  
sequence does not necessarily represent the entire amplicon.  
Sequence derived from PolyPhred was trimmed from each end of each  
unique contig until five consecutive bases exceeded a quality score  
threshold of 20, and the next 10 bases averaged a quality score of  
20 or greater. Amplicon size was estimated by agarose gel  
electrophoresis.

FEATURES  
Source

1. 847  
/organism="Sus scrofa"  
/strain="white composite, duroc, meishan, minzhu,  
fengjing, crossbreds"  
/db\_xref="taxon:9823"  
/sex="male and female"  
/clone\_lib="SCF - porcine spleen"  
/dev\_stage="adult"  
/note="Organ: spleen"  
<1..>847

STS 205 a 249 c 200 g 193 t  
ORIGIN  
Query Match 3.3%; Score 47; DB 11; Length 847;  
Best Local Similarity 100.0%; Pred. No. 4.3e-17;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 873 GCAGCACAGTGTGGAGCCACCACCATGAGAGAGCAAGTTATGGCCA 919  
|||||

Db 532 GCAGCACAGTGTGGAGCCACCACCATGAGAGAGCAAGTTATGGCCA 578  
|||||

RESULT 10  
AF069988  
LOCUS 1338 bp mRNA linear ROD 23-JUL-1998  
DEFINITION Mus musculus nitrilase 1 (Nitr1) mRNA, complete cds.  
ACCESSION AF069988  
VERSION AF069988.1 GI:3228667  
KEYWORDS

SOURCE Mus musculus  
Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1338)  
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Nitrilase and Phit homologs are encoded as fusion proteins in  
Drosophila melanogaster and Caenorhabditis elegans  
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

JOURNAL 98337986  
MEDLINE 9671749  
PUBMED 2 (bases 1 to 1338)  
REFERENCE Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
AUTHORS Direct Submission  
TITLE Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St. Philadelphia, PA 19107, USA

FEATURES  
source

1. 1338  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/map="1q21-q23"  
1..1338  
/gene="Nitr1"  
58..1029

```

/gene="Nitrilase 1"
/codon_start=1
/product="nitrilase 1"
/protein_id="AAC40185.1"
/db_xref="GI:3228668"
/translation="MLGFTIRPPHQLCTGYRLRTPVLCTQPRPTMSSSTSWELPL
VAVQGVSTPNKQNFCTCAELVQEAARLGAFLPEAFDFIARNFAETLLSLEPLN
GDLGQVSTPNKQNFCTCAELVQEAARLGAFLPEAFDFIARNFAETLLSLEPLN
GDLLGOVSQALARECGIWLISLGGFHERGQDWFQNKIYNCHVLLNSKGSVVASYRKTHL
CDVEIPGQGMRESNYTKPGGTLEPPVKTPAGKVLGAICVDMRPELSLKLQAQAGAEI
LTVPSAGSVTGPAPHEVLLPRALESQCYVIAAQCGRHETRASVGHSHVDPFWGT
VWACSEGPGLCLARIDLHFLQQRHLPVFOHRRPDLVYSLGHPLS"
BASE COUNT      347 a   335 c   330 g   326 t
ORIGIN

Query Match      2.3%; Score 33; DB 10; Length 1338;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 1375
|||||
Db 1269 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 1301

RESULT 11
BC021634
LOCUS      1365 bp   mRNA   linear   ROD 07-AUG-2002
DEFINITION Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA, complete cds.
ACCESSION BC021634
VERSION   BC021634.1 GI:18204912
KEYWORDS  MGC.
SOURCE    house mouse.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1365)
Strausberg,R.
Direct Submission
Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 18 Row: e Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6754855.
Location/Qualifiers
1..1365
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="czech ii"
/clone="MGC:13825 IMAGE:4008543"
/tissue_type="Mammary tumor metastatized to lung. MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer."
/lab_host="DH10B"
/lab_lib="NCI-CGAP_Lu30"
/label="PCMV-SPORT6"
/note="Vector: PCMV-SPORT6"
59..1030
/codon_start=1

FEATURES
source

```

```

/product="nitrilase 1"
/protein_id="AAH21634.1"
/db_xref="GI:18204913"
/db_xref="LocusID:27045"
/translation="MLGFTIRPPHQLCTGYRLRTPVLCTQPRPTMSSSTSWELPL
VAVQGVSTPNKQNFCTCAELVQEAARLGAFLPEAFDFIARNFAETLLSLEPLN
GDLGQVSTPNKQNFCTCAELVQEAARLGAFLPEAFDFIARNFAETLLSLEPLN
GDLLGOVSQALARECGIWLISLGGFHERGQDWFQNKIYNCHVLLNSKGSVVASYRKTHL
CDVEIPGQGMRESNYTKPGGTLEPPVKTPAGKVLGAICVDMRPELSLKLQAQAGAEI
LTVPSAGSVTGPAPHEVLLPRALESQCYVIAAQCGRHETRASVGHSHVDPFWGT
VWACSEGPGLCLARIDLHFLQQRHLPVFOHRRPDLVYSLGHPLS"
BASE COUNT      368 a   334 c   331 g   332 t
ORIGIN

Query Match      2.3%; Score 33; DB 10; Length 1365;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 1375
|||||
Db 1270 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 1302

RESULT 12
AF069985
LOCUS      4481 bp   DNA   linear   ROD 23-JUL-1998
DEFINITION Mus musculus nitrilase homolog 1 (Nitr1) gene, alternatively spliced product, complete cds.
ACCESSION AF069985
VERSION   AF069985.1 GI:3242979
KEYWORDS
SOURCE    Mus musculus.
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4481)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Nitrilase and Phit homologs are encoded as fusion proteins in Drosophila melanogaster and Caenorhabditis elegans
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)
98337986
PUBMED 9671749
2 (bases 1 to 4481)
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.
Direct Submission
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA
Location/Qualifiers
1..4481
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/map="1q21-q23"
541..3944
/gene="Nitr1"
/join(541..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3944)
/gene="Nitr1"
/note="nitrilase homolog 1"
/product="alternatively spliced"
541..606
/gene="Nitr1"
/note="alternatively spliced"
/join(605..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3658)
/gene="Nitr1"
/note="alternatively spliced"
/codon_start=1
/product="nitrilase homolog 1"

```

```

/protein_id="AAC40184.1"
/db_xref="GI:3242980"
/translation="MLGFTTRPQLHCTGRLRLIPVLCTQPRPTWSSSTSWELPL
VAVQVTSPTNPKQNFKTCALVOEAAELGACLAFLPEAFDFIARNPAETLLIPEPLN
GDLGQVSQLAREGGINLGLGPFHERGQDNQOKIYNCHVLLNKSNGSVASRKTHL
CDVEIPGGQMRRESNITKPGGTLEPPVKTPAGKVLGAICYDMRFPGLSLKQAQAEI
LITPSAFGSVTPPARHEVLLRARAIESQCYIAAOCGRHETRASVGHSMVDPWGT
VVARSGEGPGLCLARIDLHFLQOMRQHLVPFQHRRPDLYGLSLGHPLS"
      863..986
/gene="Niti1"
/note="alternatively spliced"
/number=2
/number=3
/number=4
/number=5
/number=6
/number=7
/number=8

```

exon

```

/number=2
/number=3
/number=4
/number=5
/number=6
/number=7
/number=8

```

exon

```

/number=3
/number=4
/number=5
/number=6
/number=7
/number=8

```

exon

```

/number=4
/number=5
/number=6
/number=7
/number=8

```

exon

```

/number=5
/number=6
/number=7
/number=8

```

exon

```

/number=6
/number=7
/number=8

```

exon

```

/number=7
/number=8

```

exon

```

/number=8

```

BASE COUNT  
ORIGIN

```

Query Match      2.3%; Score 33; DB 10; Length 4481;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1343 TGCTGAGGCTGAGCAGCAGCTGGCATTTGAAAAA 1375

Db 3898 TGCTGAGGCTGAGCAGCAGCTGGCATTTGAAAAA 3930

RESULT 13

AC105589

LOCUS

DEFINITION Rattus norvegicus clone CH230-242B2, linear HTG 13-JUL-2002

\*\*\*, 67 unordered pieces.

AC105589

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 181583)

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
 Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,

Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisegh,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,  
 Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc.help@bcm.tmc.edu  
 Project Information  
 Center project name: GNGI  
 Center clone name: CH230-242B2  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 112684 bases at least Q40  
 Consensus quality: 120012 bases at least Q30  
 Consensus quality: 127587 bases at least Q20

\*\*\*\*\*  
 NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 67 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1312: contig of 1312 bp in length  
 1313 1412: gap of unknown length  
 1413 2762: contig of 1350 bp in length  
 2763 2862: gap of unknown length  
 2863 4412: contig of 1550 bp in length  
 4413 4512: gap of unknown length  
 4513 5532: contig of 1020 bp in length  
 5533 5632: gap of unknown length  
 5633 6814: contig of 1182 bp in length

TITLE  
JOURNAL

REFERENCE  
2 (bases 1 to 181583)

Worley,K.C.

Direct Submission

JOURNAL

Submitted (09-JAN-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 181583)

Worley,K.C.

Direct Submission

JOURNAL

Submitted (13-JUL-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 12, 2002 this sequence version replaced gi:18092812.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc.help@bcm.tmc.edu

Project Information

Center project name: GNGI

Center clone name: CH230-242B2

Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 112684 bases at least Q40

Consensus quality: 120012 bases at least Q30

Consensus quality: 127587 bases at least Q20

\*\*\*\*\*

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 67 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

```
* 6815 6914: gap of unknown length
* 6915 8099: contig of 1185 bp in length
* 8100 8199: gap of unknown length
* 8200 9786: contig of 1587 bp in length
* 9787 9886: gap of unknown length
* 9887 11574: contig of 1688 bp in length
* 11575 12816: gap of unknown length
* 12817 12916: gap of unknown length
* 12917 14067: contig of 1151 bp in length
* 14068 14167: gap of unknown length
* 14168 15370: contig of 1203 bp in length
* 15371 15470: gap of unknown length
* 15471 16926: contig of 1456 bp in length
* 16927 17026: gap of unknown length
* 17027 18088: contig of 1062 bp in length
* 18089 18188: gap of unknown length
* 18189 19315: contig of 1127 bp in length
* 19316 19415: gap of unknown length
* 19416 20772: contig of 1357 bp in length
* 20773 20872: gap of unknown length
* 20873 21937: contig of 1065 bp in length
* 21938 22037: gap of unknown length
* 22038 23147: contig of 1110 bp in length
* 23148 23247: gap of unknown length
* 23248 24803: contig of 1556 bp in length
* 24804 24903: gap of unknown length
* 24904 26111: contig of 1208 bp in length
* 26112 26211: gap of unknown length
* 26212 27647: contig of 1436 bp in length
* 27648 30257: contig of 2510 bp in length
* 30258 30357: gap of unknown length
* 30358 31725: contig of 1368 bp in length
* 31726 31825: gap of unknown length
* 31826 32912: contig of 1087 bp in length
* 32913 33012: gap of unknown length
* 33013 34593: contig of 1581 bp in length
* 34594 34693: gap of unknown length
* 34694 36338: contig of 1645 bp in length
* 36339 36438: gap of unknown length
* 36439 38139: contig of 1701 bp in length
* 38140 38239: gap of unknown length
* 38240 39349: contig of 1110 bp in length
* 39350 39449: gap of unknown length
* 39450 40968: contig of 1519 bp in length
* 40969 41068: gap of unknown length
* 41069 42068: contig of 1000 bp in length
* 42069 42168: gap of unknown length
* 42169 43869: contig of 1701 bp in length
* 43870 43969: gap of unknown length
* 43970 46141: contig of 2172 bp in length
* 46142 46241: gap of unknown length
* 46242 48158: contig of 1917 bp in length
* 48159 48258: gap of unknown length
* 48259 50613: contig of 2355 bp in length
* 50614 50713: gap of unknown length
* 50714 52725: contig of 2012 bp in length
* 52726 52825: gap of unknown length
* 52826 54534: contig of 1709 bp in length
* 54535 54634: gap of unknown length
* 54635 56455: contig of 1821 bp in length
* 56456 56555: gap of unknown length
* 56556 58494: contig of 1939 bp in length
* 58495 58594: gap of unknown length
* 58595 60496: contig of 1902 bp in length
* 60497 60596: gap of unknown length
* 60597 62628: contig of 2032 bp in length
* 62629 62728: gap of unknown length
* 62729 65060: contig of 2332 bp in length
* 65061 65160: gap of unknown length
* 65161 66543: contig of 1383 bp in length
* 66544 66643: gap of unknown length
```

```
* 66644 68412: contig of 1769 bp in length
* 68413 68512: gap of unknown length
* 68513 71585: contig of 3073 bp in length
* 71586 71685: gap of unknown length
* 71686 74575: contig of 2890 bp in length
* 74576 74675: gap of unknown length
* 74676 77865: contig of 3190 bp in length
* 77866 77965: gap of unknown length
* 77966 80442: contig of 2477 bp in length
* 80443 80542: gap of unknown length
* 80543 83108: contig of 2566 bp in length
* 83109 83208: gap of unknown length
* 83209 86895: contig of 3687 bp in length
* 86896 86995: gap of unknown length
* 86996 91038: contig of 4043 bp in length
* 91039 91138: gap of unknown length
* 91139 94512: contig of 3274 bp in length
* 94513 96959: contig of 2447 bp in length
* 96960 97059: gap of unknown length
* 97060 98988: contig of 1929 bp in length
* 98989 99088: gap of unknown length
* 99089 101633: contig of 2545 bp in length
* 101634 101733: gap of unknown length
* 101734 105899: contig of 4166 bp in length

Query Match 2.3%; Score 33; DB 2; Length 181583;
Best Local Similarity 100.0%; Pred. No. 5.2e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1343 TGCTGAGGCTGAGCAGCTGGCATTGAAAAA 1375
|||||
Db 135691 TGCTGAGGCTGAGCAGCTGGCATTGAAAAA 135723
```

## RESULT 14

```
AC084821          AC084821          211772 bp      DNA      linear      ROD 19-FEB-2002
LOCUS             Mus musculus chromosome 1 clone rp23-395h6, complete sequence.
DEFINITION
ACCESSION          AC084821
VERSION            AC084821.25 GI:18702388
KEYWORDS            HTG.
SOURCE              Mus musculus.
ORGANISM            Mus musculus.
REFERENCE           1 (bases 1 to 211772)
AUTHORS             Deschamps,S., Gu,W. and Roe,B.A.
TITLE               Mus musculus BAC Clone rp23-395h6
JOURNAL             Unpublished
REFERENCE           2 (bases 1 to 211772)
AUTHORS             Deschamps,S., Gu,W. and Roe,B.A.
TITLE               Direct Submission
JOURNAL             Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
                   The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                   OK 73019, USA
REFERENCE           3 (bases 1 to 211772)
AUTHORS             Deschamps,S., Gu,W. and Roe,B.A.
TITLE               Direct Submission
JOURNAL             Submitted (19-FEB-2002) Department Of Chemistry And Biochemistry,
                   The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                   OK 73019, USA
COMMENT             On Feb 19, 2002 this sequence version replaced gi:18390260.
FEATURES             Center: Department Of Chemistry And Biochemistry
                   The University Of Oklahoma
                   Center code:UOKNOR
                   -----
                   Location/Qualifiers
                   1..211772
                   /organism="Mus musculus"
                   /db_xref="taxon:10090"
                   /chromosome="1"
```



```
/clone="rp23-395h6"
/clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT 55607 a 48743 c 49203 g 58215 t 4 others
ORIGIN

Query Match 2.3%; Score 33; DB 10; Length 211772;
Best Local Similarity 100.0%; Pred. No. 5.3e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 1375
|||||
Db 3237 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 3269

RESULT 15
AC087229/c
LOCUS AC087229 215043 bp DNA linear ROD 14-AUG-2002
DEFINITION Mus musculus chromosome 1 clone rp23-191a19 strain C57BL/6J,
complete sequence.
ACCESSION AC087229
VERSION AC087229.24 GI:22094398
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 215043)
Jiang,X., Song,L., Gu,W. and Roe,B.A.
Mus musculus Chromosome 1 BAC Clone rp23-191a19
Unpublished
2 (bases 1 to 215043)
Jiang,X., Song,L., Gu,W. and Roe,B.A.
Direct Submission
Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 215043)
Jiang,X., Song,L., Gu,W. and Roe,B.A.
Direct Submission
Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 215043)
Jiang,X., Song,L., Gu,W. and Roe,B.A.
Direct Submission
Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 215043)
Jiang,X., Song,L., Gu,W. and Roe,B.A.
Direct Submission
Submitted (13-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 215043)
Jiang,X., Song,L., Gu,W. and Roe,B.A.
Direct Submission
Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 3, 2002 this sequence version replaced gi:22038565.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----

FEATURES
Source
Location/Qualifiers
1..215043
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-191a19"

/clone_lib="RPCI - 23 Female (c57BL/6J) Mouse BAC Library"
BASE COUNT 55806 a 49340 c 51162 g 58732 t 3 others
ORIGIN

Query Match 2.3%; Score 33; DB 10; Length 215043;
Best Local Similarity 100.0%; Pred. No. 5.3e-08;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 1375
|||||
Db 87663 TGCTGAGGCTGAGCAGCACTGGCATTGAAAAA 87631

Search completed: November 30, 2002, 13:41:10
Job time : 3446 secs
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:36:49 ; Search time 253 Seconds  
(without alignments)  
12604.070 Million cell updates/sec

Title: US-09-357-675C-1

Perfect score: 1416

Sequence: 1 gccactcgtcgccctnt.....aaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1314	92.8	1416	21 AAZ46101	Nucleotide sequenc
2	1160	81.9	1382	21 AAF18257	Human prostate can
3	766	54.1	1203	20 AAX30398	DNA encoding a hum
c 4	326	23.0	377	20 ABK39058	CDNA encoding lung
5	91	6.4	505	22 AAH07371	Human CDNA clone (
6	87	6.1	592	22 AAH11324	Human CDNA clone (
c 7	87	6.1	2261	22 AAH14367	Human CDNA sequenc
c 8	87	6.1	2261	22 ABK46133	CDNA encoding colo
9	58	4.1	60	24 ABN39628	Human spliced tran

10	52	3.7	539	22	AAK61236	Human immune/haema
11	30	2.1	3300	20	AAZ11915	Human potassium ch
12	29	2.0	80	24	ABA82722	Human protective D
13	29	2.0	223	23	ABV08728	Human prostate exp
c 14	29	2.0	311	24	ABL82990	Human ovarian canc
c 15	29	2.0	384	23	ABV13242	Human prostate exp
c 16	29	2.0	406	23	ABV04073	Human prostate exp
c 17	29	2.0	416	23	ABV34361	Human prostate exp
18	29	2.0	417	23	ABV58938	Human prostate exp
19	29	2.0	528	23	ABV38626	Human prostate exp
20	29	2.0	563	20	AAK98006	Human secreted pro
c 21	29	2.0	564	24	ABN93758	Gene #256 used to
c 22	29	2.0	748	22	AAK63693	Human immune/haema
23	29	2.0	807	24	ABA82714	Human protective D
24	29	2.0	1193	13	AAQ30670	GGF28PP2.CDS. Syn
25	29	2.0	1193	15	AAQ58303	GGF-II cDNA sequen
26	29	2.0	1193	15	AAQ62849	GGF-II cDNA sequen
27	29	2.0	1193	16	AAQ74885	Putative bovine gl
28	29	2.0	1193	17	AAT48079	Bovine neuregulin
29	29	2.0	1193	17	AAT30997	Bovine glial growt
30	29	2.0	1193	17	AAT06703	Bovine glial growt
31	29	2.0	1211	22	AAH31352	Human secreted pro
32	29	2.0	1608	21	AAC95501	Human secreted pro
33	29	2.0	2317	22	AAS03917	Human secreted pro
34	29	2.0	3625	23	AAF87120	NOV9 coding sequen
c 36	28	2.0	4843	20	AAX37553	Human HLIG-1 cDNA
37	28	2.0	319	24	ABL79372	Human ovarian canc
38	28	2.0	324	23	ABV49621	Human prostate exp
39	28	2.0	326	23	ABV19858	Human prostate exp
40	28	2.0	428	23	ABV17820	Human prostate exp
c 41	28	2.0	469	23	ABV47612	Human prostate exp
c 42	28	2.0	619	22	AAI22931	Human breast canc
c 43	28	2.0	711	22	AAI97858	Human neuroblastom
44	28	2.0	941	24	ABN98591	Arabidopsis thalia
45	28	2.0	1108	13	AAQ30671	GGF28PP3.CDS. Syn
			1108	15	AAQ58304	GGF-II cDNA sequen

ALIGNMENTS

RESULT 1  
AAZ46101  
ID AAZ46101 standard; DNA: 1416 BP.

XX AC AAZ46101;

XX DT 05-MAY-2000 (first entry)

XX DE Nucleotide sequence of the coding region of NIT1 gene.

XX KW NIT1 gene: nitrilase; tumour suppressor gene; PHIT; chromosome 3p14.2;  
XX KW FRA3B; cancer; genome allele inactivation; ss.

XX OS Homo sapiens.

XX OS Mus sp.

XX OS Drosophila melanogaster.

XX OS Caenorhabditis elegans.

XX FH Key Location/Qualifiers

XX CDS 3..1415

XX FT /\*tag= a

XX FT /product= (pos: 18..20, aa: Xaa)

XX FT /product= (pos: 1179..1181, aa: Gln)

XX FT /product= (pos: 1182..1184, aa: Ala)

XX FT /note= "contains 7 internal stop codons; Xaa is an unspecified amino acid"

XX PN WO200003685-A2.

XX PD 27-JAN-2000.

XX PF 20-JUL-1999; 99WO-US16366.



KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease; ss.  
XX  
XX Homo sapiens.  
XX WO200055174-A1.  
XX  
XX PD 21-SEP-2000.  
XX  
XX PD 08-MAR-2000; 2000WO-US05988.  
XX  
XX PF 12-MAR-1999; 99US-0124270.  
XX  
XX PR (HUMA-) HUMAN GENOME SCI INC.  
XX PA (ROSE/) ROSEN C A.  
XX  
XX PI Rosen CA, Ruben SM;  
XX  
XX DR WPI; 2000-587513/55.  
XX P-PSDB; AAB57054.  
XX  
XX PT Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -  
XX  
XX PS Claim 1; Page 1124; 2338pp; English.  
XX  
XX CC AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
XX  
XX SQ Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;  
Query Match 81.9%; Score 1160; DB 21; Length 1382;  
Best Local Similarity 99.9%; Pred No. 0;  
Matches 1280; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 110 GCTGGGCTTCATCACCAGGCTCCTCACAGATTCCTCTCTGTGTGCTGGACTCCG 169  
DB 79 GCTGGGCTTCATCACCAGGCTCCTCACAGATTCCTCTCTGTGTGCTGGACTCCG 138  
QY 170 GATACCTCAACTCTAGTACTTGTGCTCAGCCAGGCCAGCCATGGCTATCTCCTC 229  
DB 139 GATACCTCAACTCTAGTACTTGTGCTCAGCCAGGCCAGCCATGGCTATCTCCTC 198  
QY 230 TTCCTCTCGGAACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 289  
DB 199' TTCCTCTCGGAACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 258  
QY 290 ACAGAACTTTAAACATGTGCTGAGCTGGTTCGAGAGGCTGCGAGACTGGGTGCTGCT 349  
DB 259 ACAGAACTTTAAACATGTGCTGAGCTGGTTCGAGAGGCTGCGAGACTGGGTGCTGCT 318  
QY 350 GGTTCCTTCCTGCTGAGGCAATTTGACTTCAATTCACCGGACCCCTGCAGACGCTACACCT 409  
DB 319 GGTTCCTTCCTGCTGAGGCAATTTGACTTCAATTCACCGGACCCCTGCAGACGCTACACCT 378  
QY 410 GTCTGAACCACTGGGGGAAACTTTTGAAGAATAACACCCAGCTTGCAGGGGAATGCG 469  
DB 379 GTCTGAACCACTGGGGGAAACTTTTGAAGAATAACACCCAGCTTGCAGGGGAATGCG 438

QY 470 ACTCTGGCTGTCTCTGGGTGGTTTCCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAA 529  
DB 439 ACTCTGGCTGTCTCTGGGTGGTTTCCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAA 498  
QY 530 AATCTACAATTTGTACGTGCTGCTGAACAGCAAAAGGGGAGTAGTGGCCACTTACAGGAA 589  
DB 499 AATCTACAATTTGTACGTGCTGCTGAACAGCAAAAGGGGAGTAGTGGCCACTTACAGGAA 558  
QY 590 GACACATCTGTGTGAGTAGAGATTCCAGGCGAGGGGCTATGTGTGAAGCAACTCTAC 649  
DB 559 GACACATCTGTGTGAGTAGAGATTCCAGGCGAGGGG-CTATGTGTGAAGCAACTCTAC 617  
QY 650 CATGCTGGGCCCCAGTCTTTGAGTCACTGTACAGCACACCAGCAGGCAAGATTGGTCTAGC 709  
DB 618 CATGCTGGGCCCCAGTCTTTGAGTCACTGTACAGCACACCAGCAGGCAAGATTGGTCTAGC 677  
QY 710 TGTCTGCTATGACATGCGGTTCCCTGAACCTCTCTGCGCATTTGGCTCAAGCTGGAGCAGA 769  
DB 678 TGTCTGCTATGACATGCGGTTCCCTGAACCTCTCTGCGCATTTGGCTCAAGCTGGAGCAGA 737  
QY 770 GATACCTTACCTATCTTTCAGCTTTTGGATCCATTACAGGCCCGCCAGCTGGGAGGTGT 829  
DB 738 GATACCTTACCTATCTTTCAGCTTTTGGATCCATTACAGGCCCGCCAGCTGGGAGGTGT 797  
QY 830 GCTGGGGGCCCCGTGTATTCGAAACCCAGTGTCTATGTAGTGGCAGCAGCAGTGTGGACG 889  
DB 798 GCTGGGGGCCCCGTGTATTCGAAACCCAGTGTCTATGTAGTGGCAGCAGCAGTGTGGACG 857  
QY 890 CCACCATGAGAGAGAGCAAGTTATTTGGCCACAGCATGTGGGTAGAGCCCTGGGGAACAGT 949  
DB 858 CCACCATGAGAGAGAGCAAGTTATTTGGCCACAGCATGTGGGTAGAGCCCTGGGGAACAGT 917  
QY 950 GTGTGCCCCCTCTCTGAGGGGCCAGGCTCTGCTTGCCTCCGAATAGACTCAACTATCT 1009  
DB 918 GTGTGCCCCCTCTCTGAGGGGCCAGGCTCTGCTTGCCTCCGAATAGACTCAACTATCT 977  
QY 1010 GCGACAGTTGCGCCGACACCTGCTGTGTCCAGCACCCAGGCTGACCTCTATGGCAA 1069  
DB 978 GCGACAGTTGCGCCGACACCTGCTGTGTCCAGCACCCAGGCTGACCTCTATGGCAA 1037  
QY 1070 TCTGGGTACCCACATGCTCTATAGACTTGTGCTGAGTTTGTAGCTTGTGCTGCTGCTGCT 1129  
DB 1038 TCTGGGTACCCACATGCTCTTAAAGACTTGTGCTGAGTTTGTAGCTTGTGCTGCTGCTGCT 1097  
QY 1130 CCACCTCTGCCACTATGAGCTAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189  
DB 1098 CCACCTCTGCCACTATGAGCTAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1157  
QY 1190 CCCCTCACTTGGAGAACCTTGACTCTCTTTGATGGAAACACAGATGGGCTGCTTGGGAAAGA 1249  
DB 1158 CCCCTCACTTGGAGAACCTTGACTCTCTTTGATGGAAACACAGATGGGCTGCTTGGGAAAGA 1217  
QY 1250 AACTTTCACCTGAGCTTTCACCTGAGCTGAGCTGAGCTGAGCTTTCAGAAAGGTGGAAATTTATA 1309  
DB 1218 AACTTTCACCTGAGCTTTCACCTGAGCTGAGCTGAGCTGAGCTTTCAGAAAGGTGGAAATTTATA 1277  
QY 1310 TAGTCACTTCTTTATTTTCATGGAAACTGAAAGTTCTGCTGAGGGCTGAGCAGCAGCTGCACTT 1369  
DB 1278 TAGTCACTTCTTTATTTTCATGGAAACTGAAAGTTCTGCTGAGGGCTGAGCAGCAGCTGCACTT 1337  
QY 1370 GAAAAATAATAATCATATAA 1390  
DB 1338 GAAAAATAATAATCATATAA 1358  
RESULT 3  
AA30398  
ID AA30398 standard; DNA; 1203 BP.  
XX  
AC AA30398;  
XX  
DT 14-MAY-1999 (first entry)  
XX

DE DNA encoding a human secreted protein.

XX Secreted protein; cancer; tumour; neurodegenerative disorder;

KW developmental abnormality; foetal deficiency; blood disorder;

KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;

KW renal disease; diabetes; inflammation; allergy; ischemic shock;

KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;

KW prostate disease; asthma; osteoporosis; arthritis; ss.

XX Homo sapiens.

XX WO9907891-A1.

PN 18-FEB-1999.

XX 04-AUG-1998; 98WO-US16235.

XX 19-AUG-1997; 97US-0056732.

PR 05-AUG-1997; 97US-0054798.

PR 05-AUG-1997; 97US-0054803.

PR 05-AUG-1997; 97US-0054804.

PR 05-AUG-1997; 97US-0054806.

PR 05-AUG-1997; 97US-0054807.

PR 05-AUG-1997; 97US-0054808.

PR 05-AUG-1997; 97US-0054809.

PR 05-AUG-1997; 97US-0055309.

PR 05-AUG-1997; 97US-0055310.

PR 05-AUG-1997; 97US-0055312.

PR 05-AUG-1997; 97US-0055386.

PR 05-AUG-1997; 97US-0055311.

PR 18-AUG-1997; 97US-0055970.

PR 19-AUG-1997; 97US-0055986.

PR 19-AUG-1997; 97US-0056365.

PR 19-AUG-1997; 97US-0056366.

PR 19-AUG-1997; 97US-0056557.

PR 19-AUG-1997; 97US-0056370.

PR 19-AUG-1997; 97US-0056371.

PR 19-AUG-1997; 97US-0056563.

PR 19-AUG-1997; 97US-0056731.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;

PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;

XX P-PSDB; AAY10877.

DR WPI: 1999-167452/14.

DR P-PSDB; AAY10877.

XX New isolated human genes encoding secreted polypeptides - useful for

PT diagnosis and treatment of pathological diseases

XX Claim 3; Page 265-266; 331pp; English.

XX The specification describes secreted proteins and their corresponding

CC polynucleotides which are useful for preventing, treating or ameliorating

CC medical conditions, e.g. by protein or gene therapy. Pathological

CC conditions can also be diagnosed by determining the amount of the

CC secreted polypeptides in a sample or by determining the presence of

CC mutations in the polynucleotides. Specific uses are described for each

CC of the products, based on which tissues they are most highly

CC expressed in, and include developing products for the diagnosis or

CC treatment of cancer, tumours, neurodegenerative disorders, developmental

CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,

CC diseases of the immune system, autoimmune diseases, hepatic and renal

CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's

CC and cognitive disorders, schizophrenia, cardiovascular disorders,

CC prostate diseases, asthma, disorders involving osteoclasts such as

CC osteoporosis, arthritis or malignancies, diseases of testes, lung or

CC thymus, digestive/endocrine disorders, infections and AIDS. The

CC polypeptides are also useful for identifying their binding partners.

XX Sequence 1203 BP: 292 A: 307 C; 325 G; 278 T: 1 other;

Query Match 54.1%; Score 766; DB 20; Length 1203;

Best Local Similarity 100.0%; Pred. No. 8.4e-305;

Matches 766; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 CTATGTGTGAAGCAACTCTTACCATGCTGGGCCAGCTCTTGAGTCACCTGTCACACAC 687

DB 161 CTATGTGTGAAGCAACTCTTACCATGCTGGGCCAGCTCTTGAGTCACCTGTCACACAC 220

QY 688 CAGCAGCAAGATTGCTGTAGCTGTCTATGACATCGGTTCCCTGAACCTCTCTCTGG 747

DB 221 CAGCAGCAAGATTGCTGTAGCTGTCTATGACATCGGTTCCCTGAACCTCTCTCTGG 280

QY 748 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTTCAGCTTTTGGATCCATTACAG 807

DB 281 CATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTTCAGCTTTTGGATCCATTACAG 340

QY 808 GCCCAGCCCACTGGAGAGTGTGCTGGGCCGCTGCTATCGAAACCCAGTCTATGTAG 867

DB 341 GCCCAGCCCACTGGAGAGTGTGCTGGGCCGCTGCTATCGAAACCCAGTCTATGTAG 400

QY 868 TGGCAGCAGCAGTGTGAGCCACCATGAGAGAGAGCAAGTTATGGCCACACATGG 927

DB 401 TGGCAGCAGCAGTGTGAGCCACCATGAGAGAGAGCAAGTTATGGCCACACATGG 460

QY 928 TGGTAGACCCCTGGGGAACAGTGTGGCCGCTGCTGAGGGGCCAGCCCTCTGCCTTG 987

DB 461 TGGTAGACCCCTGGGGAACAGTGTGGCCGCTGCTGAGGGGCCAGCCCTCTGCCTTG 520

QY 988 CCCGAATAGACCTCAACTATCTGCGACAGTGGCCGACACCTGCTGTGTTCAGCACC 1047

DB 521 CCCGAATAGACCTCAACTATCTGCGACAGTGGCCGACACCTGCTGTGTTCAGCACC 580

QY 1048 GCAGGCTGACCTCTATGCAATCTGGGTGCTGACCCACTGCTTAAGACTTGACTTCTGTGA 1107

DB 581 GCAGGCTGACCTCTATGCAATCTGGGTGCTGACCCACTGCTTAAGACTTGACTTCTGTGA 640

QY 1108 GTTTAGACCTGCCCTCCACCCACCCCTGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTG 1167

DB 641 GTTTAGACCTGCCCTCCACCCACCCCTGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTG 700

QY 1168 GAGCAGGATCAGGCACAGCTCCCTCCTTGGAGAACCTTGTGCTCTTGTATGGAACA 1227

DB 701 GAGCAGGATCAGGCACAGCTCCCTCCTTGGAGAACCTTGTGCTCTTGTATGGAACA 760

QY 1228 CAGATGGCTGCTTGGGAAAGAACTTTCACCTGAGCTTACCTGAGCTCAGCTGCACT 1287

DB 761 CAGATGGCTGCTTGGGAAAGAACTTTCACCTGAGCTTACCTGAGCTCAGCTGCACT 820

QY 1288 TTCAGAAAGTGGAAATTTTATATAGTCAATGTTTATTTTCATGGAAACTCAAGTTCGCTG 1347

DB 821 TTCAGAAAGTGGAAATTTTATATAGTCAATGTTTATTTTTCATGGAAACTCAAGTTCGCTG 880

QY 1348 AGGCTGAGCAGCAGCTGCGATTGAAATAATAATAATCAATAAGTC 1393

DB 881 AGGCTGAGCAGCAGCTGCGATTGAAATAATAATAATAATAAGTC 926

RESULT 4

ABK39058/c

ID ABK39058 standard; cDNA; 377 BP.

XX

AC ABK39058;

XX

DT 21-MAY-2002 (first entry)

XX

DE cDNA encoding lung tumour protein clone R0130.B11.

XX

KW Lung tumour; cancer; T cell; immune response stimulator;

XX cytostatic; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200204514-A2.



```
Db 19 TCTGGCTCAGACCGCCCTCCGAGTCGGACCTGCGAATGGTTTGGCTATATCTTCATG 78
Qy 80 TAGGACCTACTCCCTATCCGTCGCGCGCGG 110
Db 79 TAGGACCTACTCCCTATCCGTCGCGCGCGG 109

RESULT 6
ID AAH11324
XX AAH11324 standard; cDNA; 592 BP.
AC AAH11324;
DT 26-JUN-2001 (first entry)
DE Human cDNA clone (3'-primer) SEQ ID NO:8159.
XX Human; primer; detection: diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 3; SEQ ID 8159; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC of the present invention.
XX Sequence 592 BP; 164 A; 134 C; 138 G; 148 T; 8 other;
```

```
Query Match 6.1%; Score 87; DB 22; Length 592;
Best Local Similarity 100.0%; Pred. No. 4.5e-26;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1307 ATATAGTCAATGTTTATTTTCATGGAAACTGAAGTTCTGCTGAGGCGCTGAGCAGCACTGGC 1366
Db 1 ATATAGTCAATGTTTATTTTCATGGAAACTGAAGTTCTGCTGAGGCGCTGAGCAGCACTGGC 60

Qy 1367 ATTGAAAAATATAATAATCATAAAGTC 1393
Db 61 ATTGAAAAATATAATAATCATAAAGTC 87

RESULT 7
ID AAH14367/c
XX AAH14367 standard; cDNA; 2261 BP.
AC AAH14367;
DT 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:11772.
XX Human; primer; detection: diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 11772; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
```



CC represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.

SQ Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;  
Query Match 6.1%; Score 87; DB 22; Length 2261;  
Best Local Similarity 100.0%; Pred. No. 3.9e-26;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 ATATAGTCATTCTTTATTTTCATGGAAGTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366  
|||||  
DB 2261 ATATAGTCATTCTTTATTTTCATGGAAGTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202  
|||||  
QY 1367 ATTGAAAAATAAATAATCAATAAAGTC 1393  
|||||  
DB 2201 ATTGAAAAATAAATAATCAATAAAGTC 2175  
|||||

RESULT 8  
ABK46133/c  
ID ABK46133 standard; cDNA; 2261 BP.  
AC ABK46133;  
XX  
XX  
DT 05-JUN-2002 (first entry)  
XX  
XX cDNA encoding colon tumour protein, SEQ ID No 1684.  
XX  
XX Human; colon tumour; vaccine; colon cancer; immunogenic;  
KW immunotherapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200212328-A2.  
XX  
XX PD 14-FEB-2002.  
XX  
XX PF 31-JUL-2001; 2001WO-US24218.  
XX  
XX PR 03-AUG-2000; 2000US-223283P.  
XX  
XX PR 28-MAR-2001; 2001US-279763P.  
XX  
XX PR 29-JUN-2001; 2001US-302051P.  
XX

XX (CORI-) CORIXA CORP.  
XX  
XX King GE, Meagher MJ, Xu J, Secrist H;  
XX  
XX WPI; 2002-241739/29.  
XX  
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,  
PT for diagnosing, preventing, and treating colon cancer, and as markers  
PT for the progression of cancer -  
XX  
XX PS Claim 1; SEQ ID No 1684; 147pp; English.

XX The invention relates to polynucleotides encoding colon tumour proteins.  
XX The polynucleotides and encoded polypeptides are useful in pharmaceutical  
XX compositions, such as vaccines, for the diagnosis, prevention, and  
XX treatment of colon cancer. Polynucleotide sequences may be used as  
XX hybridisation probes or primers, and in the design and preparation of  
XX ribozyme molecules for inhibiting expression of tumour polypeptides and  
XX proteins in tumour cells. The compositions are useful for stimulating an  
XX immune response against cancer, particularly for the immunotherapy of  
XX colon cancer, and as markers for the progression of cancer.  
XX ABK4450-ABK46237 represent coding sequences of human colon tumour  
XX proteins of the invention.  
XX Note: With the exception of SEQ ID No 1 and 2, the sequence data  
XX for this patent did not form part of the printed specification but was  
XX supplied by the European Patent Office.

XX SQ Sequence 2261 BP; 543 A; 594 C; 541 G; 583 T; 0 other;  
Query Match 6.1%; Score 87; DB 24; Length 2261;

Best Local Similarity 100.0%; Pred. No. 3.9e-26;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1307 ATATAGTCATTCTTTATTTTCATGGAAGTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 1366  
|||||  
DB 2261 ATATAGTCATTCTTTATTTTCATGGAAGTGAAGTTCTGCTGAGGGCTGAGCAGCACTGGC 2202  
|||||  
QY 1367 ATTGAAAAATAAATAATCAATAAAGTC 1393  
|||||  
DB 2201 ATTGAAAAATAAATAATCAATAAAGTC 2175  
|||||

RESULT 9  
ABN39628  
ID ABN39628 standard; DNA; 60 BP.  
XX  
XX AC ABN39628;  
XX  
XX DT 15-JUL-2002 (first entry)  
XX  
XX Human spliced transcript detection oligonucleotide SEQ ID NO:12376.  
XX  
XX DE Human; mouse; rat; splice transcript; detection; RNA transcript;  
XX KW splice variant; transcriptome; oligonucleotide library; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200210449-A2.  
XX  
XX PD 07-FEB-2002.  
XX  
XX PF 20-JUL-2001; 2001WO-IB01903.  
XX  
XX PR 28-JUL-2000; 2000US-221607P.  
XX  
XX PR 02-MAY-2001; 2001US-287724P.  
XX  
XX PA (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
XX WPI; 2002-257383/30.  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes -  
XX  
XX PS Example 1; SEQ ID 12376; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting  
XX messenger RNAs that populate a (sub-)transcriptome, where the  
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
XX transcription units that populate a genome. The library comprises  
XX several oligonucleotides, each capable of hybridising selectively to a  
XX set of messenger RNAs transcribed from a given transcription unit of  
XX the genome, which encodes one or more messenger RNA splice variants.  
XX The oligonucleotide libraries are useful for detecting mRNAs from a  
XX biological sample, in expression profiling studies, in qualitatively or  
XX quantitatively characterising the corresponding transcriptome, and in  
XX detecting RNA transcripts and splice variants of human or animal  
XX transcriptomes. The libraries may also be used as specialised mini  
XX libraries to detect transcripts of a sub-transcriptome under a  
XX particular biological or pathological state, and so allowing the  
XX detection of tissue- or pathology-specific genes such as those genes  
XX only expressed in specific tissue under a specific pathological  
XX condition; to detect developmental specific genes; and to detect RNA  
XX transcripts and splice variants of a transcriptome of a patient suffering  
XX from a particular disorder. ABN27253 to ABN59589 represent  
XX oligonucleotide sequences from rats, humans and mice, which are used in  
XX the exemplification of the present invention.  
XX N.B. The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

```
XX SQ Sequence 60 BP; 9 A; 20 C; 14 G; 17 T; 0 other;
Query Match 4.1%; Score 58; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 4.5e-14;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 TCTGGCTCCAGACCGCCCTCCGGATCGGACCTGGCAATGGTTTGGCTATATCTTCA 77
|||||
DB 3 TCTGGCTCCAGACCGCCCTCCGGATCGGACCTGGCAATGGTTTGGCTATATCTTCA 60

RESULT 10
AAK61236
ID AAK61236 standard; cDNA; 539 BP.
XX AC AAK61236;
XX DT
XX 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6296.
XX DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
```

```
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-483426/52.
XX P-PSDB; AAM88455.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Claim 1; SEQ ID NO 6296; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients' own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;
XX
XX Query Match 3.7%; Score 52; DB 22; Length 539;
XX Best Local Similarity 100.0%; Pred. No. 1e-11;
XX Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 109 GGCTGGGCTTCATCACAGCGCTCCTCACAGATTCCTGTCCTTCGTGTC 160
DB 371 GGCTGGGCTTCATCACAGCGCTCCTCACAGATTCCTGTCCTTCGTGTC 422
XX
RESULT 11
AAZ11915
ID AAZ11915 standard; cDNA; 3300 BP.
XX
XX AAZ11915;
XX
XX 30-NOV-1999 (first entry)
XX
```

```
XX Human potassium channel K+Hnov59 cDNA.
XX
XX Potassium channel; ataxia; arrhythmia; epilepsy; Bartter's syndrome;
XX cardiovascular disorder; CNS disorder; renal disorder; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 50..1285
XX /*tag= a
XX /product= "Human K+Hnov59 potassium channel"
XX
XX WO9943696-A1.
XX
XX 02-SEP-1999.
XX
XX 22-FEB-1999; 99WO-US03826.
XX
XX 19-JAN-1999; 99US-0116448.
XX 25-FEB-1998; 98US-0076687.
XX 07-AUG-1998; 98US-0095836.
XX
XX (AXYS-) AXYS PHARM INC.
XX
XX Curran ME, Hu P, Miller AP, Rutter M, Wang J;
XX
XX WPI: 1999-527591/44.
XX P-PSDB; AAY34133.
XX
XX New nucleic acids encoding mammalian K+Hnov potassium channel
XX proteins, useful for the diagnosis and treatment of episodic ataxia
XX with myokymia, cardiac arrhythmia, epilepsy and Bartter's syndrome
XX
XX Claim 4; Page 102-104; 112pp; English.
XX
XX This sequence represents human potassium channel K+Hnov59 cDNA.
XX K+Hnov proteins have a high degree of homology to known potassium
XX channels and may be alpha subunits, which form the functional channel,
XX or accessory subunits that act to modulate the channel activity. K+Hnov59
XX is a 4 transmembrane domain, 2 pore domain potassium channel. The gene
XX is located on chromosome 19, determined via PCR chromosomal
XX localisation using primers AAZ11939 and AAZ11940. K+Hnov cDNAs
XX were isolated by extension of expressed sequence tags (ESTs) which were
XX related but not identical to known human potassium channels. Potential
XX polymorphisms detected as sequence variants between multiple
XX independent clones. Potassium channels have critical roles in various
XX cell types and biochemical pathways. Defective potassium channels are
XX known to cause four human diseases: episodic ataxia with myokymia;
XX cardiac arrhythmia (long QT syndrome); epilepsy; and Bartter's syndrome.
XX As potassium channels are critical components of virtually all cells,
XX it is likely that abnormal potassium channels are also implicated in
XX certain renal, cardiovascular and central nervous system (CNS)
XX disorders. Nucleotides encoding K+Hnov proteins may be used for
XX identifying homologous or related proteins and the DNA sequences encoding
XX them. They may be used to produce compositions that modulate the
XX expression and function of the K+Hnov protein and in studying the
XX biochemical pathways associated with it. They may also be used for the
XX recombinant production of K+Hnov protein in fermentation cultures.
XX Additionally, such nucleotides may be used in gene therapy protocols for
XX the treatment of diseases associated with abnormal potassium channels.
XX
XX Sequence 3300 BP; 997 A; 629 C; 680 G; 994 T; 0 other;
XX
XX Query Match 2.1%; Score 30; DB 20; Length 3300;
XX Best Local Similarity 100.0%; Pred. No. 0.0093;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1387 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
DB 3238 TAAAGTCAAAAAAAAAAAAAAAAAAAAAA 3267
XX
```

```
RESULT 12
ABA82722
ID ABA82722 standard; DNA; 80 BP.
XX
XX
AC ABA82722;
XX
DT 07-FEB-2002 (first entry)
XX
DE Human protective DNA sequence CNI-00735 fragment #8.
XX
KW Human; protective sequence; cell death; cancer; autoimmune disease;
KW neurological disorder; stroke; cytostatic; neuroprotective; gene therapy;
KW ds.
XX
XX Homo sapiens.
XX
XX WO200176457-A2.
XX
XX 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-US11663.
XX
XX 11-APR-2000; 2000US-0547735.
XX
XX (COGE-) COCENT NEUROSCIENCE INC.
XX
XX Thomas MB, Portbury SD, Puranam K, Katz LC, Lo DC, Barney S;
XX
XX WPI; 2002-025874/03.
XX
XX P-PSDB; ABB44640.
XX
XX New protective sequences and their products, useful for diagnosing and
XX treating diseases involving cell death, including neurological
XX disorders e.g. stroke and for identifying modulators of expression of
XX the protective sequences -
XX
XX Claim 2; Fig 5; 283pp; English.
XX
XX The present invention relates to protective sequence proteins
XX (ABA4624-ABBA4830) and their coding sequences (ABA82701-ABA82937).
XX The sequences, when introduced into a cell either predisposed to undergo
XX cell death or in the process of undergoing cell death, prevent, delay or
XX rescue the cell from death, hence, these sequences are named "protective
XX sequences". The sequences are useful for treating and/or ameliorating
XX cancer, autoimmune diseases and neurological disorders e.g. stroke.
XX Further examples of diseases which may be treated by the present
XX invention are given in the specification.
XX
XX Sequence 80 BP; 40 A; 10 C; 15 G; 15 T; 0 other;
XX
XX Query Match 2.0%; Score 29; DB 24; Length 80;
XX Best Local Similarity 100.0%; Pred. No. 0.035;
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1388 AAAGTCAAAAA AAAAAAAAAAAAAAAAAAAAAA 1416
Db 39 AAAGTCAAAAA AAAAAAAAAAAAAAAAAAAAAA 67
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

RESULT 13
ABV08728
ID ABV08728 standard; cDNA; 223 BP.
XX
XX
AC ABV08728;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 8719.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
```

```
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX 16-MAR-2000; 2000US-189862P.
XX 25-MAY-2000; 2000US-207454P.
XX 09-JUN-2000; 2000US-211314P.
XX 18-JUL-2000; 2000US-219007P.
XX 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 1376; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for:
XX (a) assessing whether a patient is afflicted with prostate cancer;
XX (b) monitoring the progression of prostate cancer in a patient;
XX (c) assessing the efficacy of a test compound to inhibit prostate
XX cancer in a patient;
XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound;
XX (g) determining whether prostate cancer has metastasized in a patient;
XX (h) assessing the aggressiveness or indolence of prostate cancer in a
XX patient;
XX (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 223 BP; 76 A; 44 C; 64 G; 38 T; 1 other;
XX
XX Query Match 2.0%; Score 29; DB 23; Length 223;
XX Best Local Similarity 100.0%; Pred. No. 0.031;
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1388 AAAGTCAAAAA AAAAAAAAAAAAAAAAAAAAAA 1416
Db 188 AAAGTCAAAAA AAAAAAAAAAAAAAAAAAAAAA 216
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

RESULT 14
ABL82990/c
ID ABL82990 standard; cDNA; 311 BP.
XX
XX
AC ABL82990;
XX
XX 17-MAY-2002 (first entry)
XX
XX Human ovarian cancer related cDNA clone SEQ ID NO:5968.
XX
XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200192581-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US17756.
XX
XX
```

```
PR 26-MAY-2000; 2000US-207484P.
XX (CORI-) CORIXA CORP.
PA
XX
PI Algate PA, Harlocker SL, Jones R;
XX
DR WPI; 2002-122075/16.
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide -
XX
XX Claim 1; SEQ ID 5968; 489pp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
CC ABU87934, (III) encoding (II) having a sequence (S2), a T cell
CC population of (II), or antigen presenting cells that express (II).
CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
CC (S1) can be used for detecting ovarian cancer in a patient's biological
CC sample preferably serum or ovarian tissue. The method comprises
CC contacting a biological sample from a patient with (IV), detecting the
CC amount of polynucleotide hybridising to (IV) and comparing the amount to
CC a predetermined cutoff value and thereby detecting ovarian cancer in the
CC patient, where the amount of polynucleotide hybridising to (IV) is
CC detected preferably by polymerase chain reaction (PCR). (I) comprising
CC (III) and/or (II) is useful for stimulating and/or expanding T cells
CC specific for an ovarian tumour protein comprising contacting T cells
CC with (III) or (II). (III) is useful in design and preparation of
CC ribozyme molecules for inhibiting expression of the tumour polypeptides
CC and proteins in tumour cells; and to isolate a full length gene from a
CC suitable library e.g., a tumour cDNA library using well known
CC techniques.
XX
XX Sequence 311 BP; 96 A; 42 C; 45 G; 128 T; 0 other;
SQ
Query Match 2.0%; Score 29; DB 24; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 40 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 12

RESULT 15
ABV13242/c
ID ABV13242 standard; cDNA; 384 BP.
XX
XX AC ABV13242;
XX
XX DT 13-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 13233.
XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200160860-A2.
XX
XX PD 23-AUG-2001.
XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX
XX PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 2192; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 384 BP; 136 A; 69 C; 73 G; 106 T; 0 other;
SQ
Query Match 2.0%; Score 29; DB 23; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 42 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 14

Search completed: November 30, 2002, 12:43:36
Job time : 267 secs
```





```
;
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036.555B
; FILING DATE: 24-MAR-1993
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
;
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-036-555B-134
;
; Query Match 2.0%; Score 29; DB 1; Length 1193;
; Best Local Similarity 100.0%; Pred. No. 0.0027;
; Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
; Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152
;
; RESULT 3
; US-08-469-569-134
; Sequence 134, Application US/08469569
; Patent No. 5606032
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
```

```
;
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,569
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
;
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
;
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-469-569-134
;
; Query Match 2.0%; Score 29; DB 1; Length 1193;
; Best Local Similarity 100.0%; Pred. No. 0.0027;
; Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
; Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152
;
; RESULT 4
; US-08-249-322A-134
; Sequence 134, Application US/08249322A
; Patent No. 5716930
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; PREPARATION AND USE
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
```



```
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/249,322A
; FILING DATE: 26-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 250.4
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-249-322A-134

Query Match 2.0%; Score 29; DB 1; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
| | | | | | | | | | | | | | | | | | | |
Db 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 5
US-08-469-526A-134
; Sequence 134, Application US/08469526A
; Patent No. 5792849
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew
; APPLICANT: Stroobant, Paul
; APPLICANT: Minghetti, Luisa
; APPLICANT: Waterfield, Michael
; APPLICANT: Marchionni, Mark
; APPLICANT: Hiles, Ian
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 187
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
```

```
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,526A
; FILING DATE: 06 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 03-JUN-1992
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bleker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 04585/00200A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-469-526A-134

Query Match 2.0%; Score 29; DB 1; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
| | | | | | | | | | | | | | | | | | | |
Db 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 6
US-08-734-591A-134
; Sequence 134, Application US/08734591A
; Patent No. 5854220
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew
; APPLICANT: Stroobant, Paul
; APPLICANT: Minghetti, Luisa
; APPLICANT: Waterfield, Michael
; APPLICANT: Hiles, Ian
; APPLICANT: Marchionni, Mark
; APPLICANT: Chen, Mario
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 187
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible Pentium
; OPERATING SYSTEM: Windows95
; SOFTWARE: WordPerfect (Version 7.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,591A
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 536
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/470,335
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 03-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 91 07566.3
; FILING DATE: 10-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bleker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 04585/00200P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 428-0200
; TELEFAX: (617) 428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-734-591A-134

Query Match          2.0%; Score 29; DB 2; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
    |||||
Db 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 7
US-08-469-660-134
; Sequence 134, Application US/08469660
; Patent No. 5876973
; GENERAL INFORMATION:
; APPLICANT: Gwynne, David I.; Marchionni, Mark;
; APPLICANT: McBurney, Robert N.
; TITLE OF INVENTION: INHIBITORS OF CELL PROLIFERATION,
; TITLE OF INVENTION: THEIR PREPARATION AND USE
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; ZIP: 02111-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469.660
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/011.396
```

```
; FILING DATE: 29-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/984,085
; FILING DATE: 01-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/951,747
; FILING DATE: 25-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/927,337
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04585/017004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: 200154
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-469-660-134

Query Match          2.0%; Score 29; DB 2; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
    |||||
Db 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 8
US-08-341-018-3
; Sequence 3, Application US/08341018A
; Patent No. 6087323
; GENERAL INFORMATION:
; APPLICANT: Gwynne, David I.
; APPLICANT: Mahanthappa, Nagesh K.
; APPLICANT: Marchionni, Mark A.
; APPLICANT: Birmingham-McDonogh, Olivia
; APPLICANT: Goldin, Stanley M.
; APPLICANT: McBurney, Robert N.
; TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF
; TITLE OF INVENTION: CELLULAR COMMUNICATION
; FILE REFERENCE: 04585/041001
; CURRENT APPLICATION NUMBER: US/08/341.018A
; CURRENT FILING DATE: 1994-11-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1193
; TYPE: DNA
; ORGANISM: Bos taurus
; US-08-341-018-3

Query Match          2.0%; Score 29; DB 3; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
    |||||
Db 1124 AAAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 9
US-08-470-335-134
; Sequence 134, Application US/08470335F
; Patent No. 6147190
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
```

APPLICANT: STROOBANT, PAUL  
APPLICANT: MINGHETTI, LUISA  
APPLICANT: WATERFIELD, MICHAEL  
APPLICANT: MARCHIONNI, MARK  
APPLICANT: CHEN, MARIO S.  
APPLICANT: HILES, IAN  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
FILE REFERENCE: 04585/00200B  
CURRENT APPLICATION NUMBER: US/08/470,335F  
CURRENT FILING DATE: 1995-06-06  
EARLIER APPLICATION NUMBER: 08/036,555  
EARLIER FILING DATE: 1993-03-24  
NUMBER OF SEQ ID NOS: 252  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 134  
LENGTH: 1193  
TYPE: DNA  
ORGANISM: Bos taurus  
US-08-470-335-134

Query Match 2.0%; Score 29; DB 3; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCACAAAAA 1416  
DB 1124 AAGTCACAAAAA 1152

RESULT 10  
US-08-735-021-134  
Sequence 134, Application US/08735021B  
Patent No. 6194377  
GENERAL INFORMATION:  
APPLICANT: GOODEARL, ANDREW  
APPLICANT: STROOBANT, PAUL  
APPLICANT: MINGHETTI, LUISA  
APPLICANT: WATERFIELD, MICHAEL  
APPLICANT: MARCHIONNI, MARK  
APPLICANT: CHEN, MARIO S.  
APPLICANT: HILES, IAN  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
FILE REFERENCE: 04585/00200L  
CURRENT APPLICATION NUMBER: US/08/735,021B  
CURRENT FILING DATE: 1996-10-22  
EARLIER APPLICATION NUMBER: 08/472,065  
EARLIER FILING DATE: 1995-06-06  
EARLIER APPLICATION NUMBER: 08/036,555  
EARLIER FILING DATE: 1993-03-24  
EARLIER APPLICATION NUMBER: 07/965,173  
EARLIER FILING DATE: 1992-10-23  
EARLIER APPLICATION NUMBER: 07/940,389  
EARLIER FILING DATE: 1992-09-03  
EARLIER APPLICATION NUMBER: 07/907,138  
EARLIER FILING DATE: 1992-06-30  
EARLIER APPLICATION NUMBER: 07/863,703  
EARLIER FILING DATE: 1992-04-03  
NUMBER OF SEQ ID NOS: 192  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 134  
LENGTH: 1193  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (8)...(796)  
US-08-735-021-134

Query Match 2.0%; Score 29; DB 4; Length 1193;  
Best Local Similarity 100.0%; Pred. No. 0.0027;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCACAAAAA 1416  
DB 1124 AAGTCACAAAAA 1152

RESULT 11  
US-08-734-664A-134  
Sequence 134, Application US/08734664A  
Patent No. 6204241  
GENERAL INFORMATION:  
APPLICANT: GOODEARL, ANDREW  
APPLICANT: STROOBANT, PAUL  
APPLICANT: MINGHETTI, LUISA  
APPLICANT: WATERFIELD, MICHAEL  
APPLICANT: MARCHIONNI, MARK  
APPLICANT: CHEN, MARIO  
APPLICANT: HILES, IAN  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
FILE REFERENCE: 04585/00200B  
CURRENT APPLICATION NUMBER: US/08/734,664A  
CURRENT FILING DATE: 1995-06-06  
EARLIER APPLICATION NUMBER: 08/036,555  
EARLIER FILING DATE: 1993-03-24  
NUMBER OF SEQ ID NOS: 252  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 134  
LENGTH: 1193  
TYPE: DNA  
ORGANISM: Bos taurus  
US-08-734-664A-134

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible Pentium  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/734,664A  
FILING DATE: 22-OCT-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/249,322  
FILING DATE: 26-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/036,555  
FILING DATE: 24-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: UK 91 07566.3  
FILING DATE: 10-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bleker-Brady, Kristina  
REGISTRATION NUMBER: 39,109  
REFERENCE/DOCKET NUMBER: 04585/00200J  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 428-0200  
TELEFAX: (617) 428-7045  
TELEX:

INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1193  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-734-664A-134

PCI-US95-06846A-134

```

; Sequence 134, Application PC/TUS9506846A
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew David; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06846A
; FILING DATE: 25-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/249,322
; FILING DATE: 26-MAY-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Norman D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5250.5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US95-06846A-134

Query Match 2.0%; Score 29; DB 5; Length 1193;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
Db 1124 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1152

```

Search completed: November 30, 2002, 14:02:14  
Job time : 69 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:43:44 ; Search time 64 Seconds

(without alignments)  
8520.679 Million cell updates/sec

Title: US-09-357-675C-1

Perfect score: 1416

Sequence: 1 gccacactgctgcgcctnt.....aaaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 341543 seqs, 19255720 residues

Word size : 0

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1160	81.9	1382	10	US-09-925-300-692
2	326	23.0	377	9	US-09-736-457-1096
3	326	23.0	377	9	US-09-902-941-1096
4	87	6.1	2261	10	US-09-920-300A-1684
5	87	6.1	2261	12	US-10-033-528-1684
6	32	2.3	206	10	US-09-783-590-11058
7	29	2.0	311	10	US-09-867-701-5968
8	29	2.0	564	10	US-09-880-107-256
9	28	2.0	280	10	US-09-960-352-12666
10	28	2.0	319	10	US-09-867-701-2350
11	28	2.0	941	10	US-09-770-445-359
12	28	2.0	1505	10	US-09-925-301-142
13	28	2.0	3238	10	US-09-874-628-5
14	28	2.0	3592	10	US-09-777-745-1
15	28	2.0	3627	10	US-09-777-745-5
16	28	2.0	7878	10	US-09-962-436-562
17	27	1.9	426	10	US-09-777-564-451
18	27	1.9	639	9	US-09-950-933A-12
19	27	1.9	915	10	US-09-925-301-244

C	20	27	1.9	989	10	US-09-272-162-1	Sequence 1, Appli
	21	27	1.9	1169	10	US-09-804-156-10	Sequence 10, Appl
	22	27	1.9	1334	10	US-09-789-561-30	Sequence 30, Appl
	23	27	1.9	1575	10	US-09-833-381-1159	Sequence 1159, Ap
	24	27	1.9	1689	9	US-09-989-919-71	Sequence 71, Appl
	25	27	1.9	1969	10	US-09-070-844-7	Sequence 7, Appli
	26	27	1.9	2096	10	US-09-070-844-19	Sequence 19, Appl
	27	27	1.9	2099	10	US-09-070-844-3	Sequence 3, Appli
	28	27	1.9	2137	10	US-09-070-844-18	Sequence 18, Appl
	29	27	1.9	2140	10	US-09-070-844-1	Sequence 1, Appli
	30	27	1.9	3441	9	US-09-944-413-6	Sequence 6, Appli
	31	27	1.9	3441	9	US-09-944-403-6	Sequence 6, Appli
	32	27	1.9	3441	9	US-09-944-896-6	Sequence 6, Appli
	33	27	1.9	3441	9	US-09-944-944-6	Sequence 6, Appli
	34	27	1.9	3441	10	US-09-866-028-6	Sequence 6, Appli
	35	27	1.9	3441	10	US-09-944-449-6	Sequence 6, Appli
	36	27	1.9	3441	10	US-09-944-457-6	Sequence 6, Appli
	37	27	1.9	3441	10	US-09-945-587-6	Sequence 6, Appli
	38	27	1.9	3441	10	US-09-945-015-6	Sequence 6, Appli
	39	27	1.9	3441	10	US-09-944-396-6	Sequence 6, Appli
	40	27	1.9	3441	10	US-09-944-097-6	Sequence 6, Appli
	41	27	1.9	3441	10	US-09-944-432-6	Sequence 6, Appli
	42	27	1.9	3441	10	US-09-943-762-6	Sequence 6, Appli
	43	27	1.9	3441	10	US-09-944-654-6	Sequence 6, Appli
	44	27	1.9	3441	10	US-09-943-851A-6	Sequence 6, Appli
	45	27	1.9	3547	9	US-10-044-716-7	GENERAL INFORMATI

ALIGNMENTS

RESULT 1  
US-09-925-300-692  
; Sequence 692, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Steve Ruben  
; APPLICANT: Craig Rosen,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; PRIORITY FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 692  
; LENGTH: 1382  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-300-692

Query Match	81.9%;	Score 1160;	DB 10;	Length 1382;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1280;	Conservative	0;	Mismatches	0; Indels 1; Gaps 1;
Qy	110	GCTGGCTTCATCACACAGCCCTCCTCACAGATTCCTGTCTGTCTGTGGACTCGG	169	
Db	79	GCTGGCTTCATCACACAGCCCTCCTCACAGATTCCTGTCTGTCTGTGGACTCGG	138	
Qy	170	GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGCCAGCCATGGCTATCTCTC	229	
Db	139	GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGCCAGCCATGGCTATCTCTC	198	
Qy	230	TTCTCTCTCGGAACATGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	289	
Db	199	TTCTCTCTCGGAACATGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	258	
Qy	290	ACAGAACTTTAAACATGTGCTGAGCTGGTTCGAGAGGTGCCAGACTGGGTGCCTGCT	349	
Db	259	ACAGAACTTTAAACATGTGCTGAGCTGGTTCGAGAGGTGCCAGACTGGGTGCCTGCT	318	





```

Query Match          6.1%; Score 87; DB 10; Length 2261;
Best Local Similarity 100.0%; Pred. No. 5e-26;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1307 ATATAGTCATTGTTTATTTTCATGGAAACTGAAGTTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
      |||||||
DB 2261 ATATAGTCATTGTTTATTTTCATGGAACTGAAGTTTCTGCTGAGGGCTGAGCAGCACTGGC 2202
      |||||||

QY 1367 ATTGAAAAATATAATAATCAATAAGTC 1393
      |||||||
DB 2201 ATTGAAAAATATAATAATCAATAAGTC 2175
      |||||||

RESULT 5
US-10-033-528-1684/c
; Sequence 1684, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033.528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1684
; LENGTH: 2261
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1684

Query Match          6.1%; Score 87; DB 12; Length 2261;
Best Local Similarity 100.0%; Pred. No. 5e-26;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1307 ATATAGTCATTGTTTATTTTCATGGAACTGAAGTTTCTGCTGAGGGCTGAGCAGCACTGGC 1366
      |||||||
DB 2261 ATATAGTCATTGTTTATTTTCATGGAACTGAAGTTTCTGCTGAGGGCTGAGCAGCACTGGC 2202
      |||||||

QY 1367 ATTGAAAAATATAATAATCAATAAGTC 1393
      |||||||
DB 2201 ATTGAAAAATATAATAATCAATAAGTC 2175
      |||||||

RESULT 6
US-09-783-590-11058
; Sequence 11058, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783.590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11058
; LENGTH: 206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

NAME/KEY: misc feature  
LOCATION: (5)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (47)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (52)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (60)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (76)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (149)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (164)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (172)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (187)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (190)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (198)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (204)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-11058

Query Match 2.3%; Score 32; DB 10; Length 206;  
Best Local Similarity 100.0%; Pred. No. 0.00064;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 GTAGGACCTACTCCCTATCCCTCGCGCGG 110  
Db 8 GTAGGACCTACTCCCTATCCCTCGCGCGG 39

RESULT 7  
US-09-867-701-5968/c  
; Sequence 5968, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5968  
; LENGTH: 311  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-5968

Query Match 2.0%; Score 29; DB 10; Length 311;  
Best Local Similarity 100.0%; Pred. No. 0.0098;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416

Db 40 AAGTCAAAAAAAAAAAAAAAAAAAAAA 12

RESULT 8  
US-09-880-107-256/c  
; Sequence 256, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 256  
; LENGTH: 564  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AAL27741  
US-09-880-107-256

Query Match 2.0%; Score 29; DB 10; Length 564;  
Best Local Similarity 100.0%; Pred. No. 0.0093;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
Db 43 AAGTCAAAAAAAAAAAAAAAAAAAAAA 15

RESULT 9  
US-09-960-352-12666  
; Sequence 12666, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 12666  
; LENGTH: 280  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 54-LIB3058-028-Q1-K1-F6  
US-09-960-352-12666

Query Match 2.0%; Score 28; DB 10; Length 280;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
Db 200 AAGTCAAAAAAAAAAAAAAAAAAAAAA 227

RESULT 10  
US-09-867-701-2350/c

```
; Sequence 2350, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2350
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(319)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-2350

Query Match          2.0%; Score 28; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
DB 43 AAGTCAAAAAAAAAAAAAAAAAAAAAA 16

RESULT 11
US-09-770-445-359
; Sequence 359, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 359
; LENGTH: 941
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(941)
; OTHER INFORMATION: n = A,T,C or G
```

```
US-09-770-445-359

Query Match          2.0%; Score 28; DB 10; Length 941;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
DB 912 AAGTCAAAAAAAAAAAAAAAAAAAAAA 939

RESULT 12
US-09-925-301-142
; Sequence 142, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 142
; LENGTH: 1505
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1493)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1499)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1500)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-142

Query Match          2.0%; Score 28; DB 10; Length 1505;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416
DB 1432 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1459

RESULT 13
US-09-874-628-5
; Sequence 5, Application US/09874628
; Patent No. US20020137133A1
; GENERAL INFORMATION:
; APPLICANT: WOZNEY, John
; APPLICANT: CELESTE, Anthony J.
; APPLICANT: THIES, R. Scott
; APPLICANT: YAMAJI, No. US20020137133A1ofu
; TITLE OF INVENTION: RECEPTOR PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute Inc.- Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/09/874,628  
FILING DATE: 05-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/123,934  
FILING DATE: 17-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: LAZAR, Steven R  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5203  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 876 1170  
TELEFAX: 617 876 5851  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3238 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
IMMEDIATE SOURCE:  
CLONE: CFK1-10a  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 474..2000  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-874-628-5

Query Match 2.0%; Score 28; DB 10; Length 3238;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 3155 AAGTCAAAAAAAAAAAAAAAAAAAAAA 3182

RESULT 14  
US-09-777-745-1  
; Sequence 1, Application US/09777745  
; Patent No. US20010021702A1  
; GENERAL INFORMATION:  
; APPLICANT: CALMELS, THIERRY PAUL GERARD  
; APPLICANT: SOUCHET, MICHEL LOUIS  
; APPLICANT: LEGER, ISABELLE MARIE  
; APPLICANT: TOURTELIER, LAURENCE NATHALIE PATRICIA  
; APPLICANT: BRIL, ANTOINE MICHEL ALAIN  
; APPLICANT: DOE, TRUDY RACHEL  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GH-30321-C1  
; CURRENT APPLICATION NUMBER: US/09/777.745  
; CURRENT FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: EP 98400014.1  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: EP 98400015.8  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: UK 9710910.2  
; PRIOR FILING DATE: 1997-05-27  
; PRIOR APPLICATION NUMBER: 05/082,271  
; PRIOR FILING DATE: 1998-05-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; TYPE: DNA  
; LENGTH: 3592  
; ORGANISM: HOMO SAPIENS  
US-09-777-745-1

Query Match 2.0%; Score 28; DB 10; Length 3592;  
Best Local Similarity 100.0%; Pred. No. 0.02;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 3559 AAGTCAAAAAAAAAAAAAAAAAAAAAA 3586  
RESULT 15  
US-09-777-745-5  
; Sequence 5, Application US/09777745  
; Patent No. US20010021702A1  
; GENERAL INFORMATION:  
; APPLICANT: CALMELS, THIERRY PAUL GERARD  
; APPLICANT: SOUCHET, MICHEL LOUIS  
; APPLICANT: LEGER, ISABELLE MARIE  
; APPLICANT: TOURTELIER, LAURENCE NATHALIE PATRICIA  
; APPLICANT: BRIL, ANTOINE MICHEL ALAIN  
; APPLICANT: DOE, TRUDY RACHEL  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GH-30321-C1  
; CURRENT APPLICATION NUMBER: US/09/777.745  
; CURRENT FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: EP 98400014.1  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: EP 98400015.8  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: UK 9710910.2  
; PRIOR FILING DATE: 1997-05-27  
; PRIOR APPLICATION NUMBER: 09/082,271  
; PRIOR FILING DATE: 1998-05-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 3627  
; TYPE: DNA  
; ORGANISM: HOMO SAPIENS  
US-09-777-745-5

Query Match 2.0%; Score 28; DB 10; Length 3627;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1389 AAGTCAAAAAAAAAAAAAAAAAAAAAA 1416  
|||||  
Db 3594 AAGTCAAAAAAAAAAAAAAAAAAAAAA 3621

Search completed: November 30, 2002, 14:03:47  
Job time : 83 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2002, 12:38:19 ; Search time 1858 Seconds  
(without alignments)  
12342.739 Million cell updates/sec

Title: US-09-357-675C-1  
Perfect score: 1416  
Sequence: 1 gccacactgctggcgctct.....aaaaaaaaaaaaaaaaaaaa 1416

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.\*

```
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_nus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	844	59.6	844	9 AL529152	AL529152 AL529152
2	813	57.4	890	9 AL520768	AL520768 AL520768
3	798	56.4	943	14 BM804704	BM804704 AGENCOURT
4	691	48.8	960	9 AL522373	AL522373 AL522373
5	677	47.8	782	14 B0441752	B0441752 AGENCOURT
6	675	47.7	1048	14 B0059007	B0059007 AGENCOURT

c 7	660	46.6	677	14	BQ581760	BQ581760 ill12906.x
c 8	645	45.6	659	13	BM666784	UI-E-C11-
c 9	625	44.1	625	13	BI711300	BM666784 UI-E-C11-
c 10	610	43.1	729	14	BM975958	UI-CF-EN1
c 11	606	42.8	613	14	BM690873	UI-E-C11-
c 12	604	42.7	846	13	BI769604	BM690873 UI-E-C11-
c 13	589	41.6	750	14	BM979227	BI769604 603054985
c 14	585	41.3	992	9	AL520767	BM979227 UI-CF-DU1
c 15	584	41.2	586	9	AI797259	AL520767 AL520767
c 16	577	40.7	600	14	BM726885	AI797259 we86e02.x
c 17	573	40.5	1150	14	BM925975	BM726885 UI-E-EJ0-
c 18	572	40.4	576	13	BI711712	BM925975 AGENCOURT
c 19	571	40.3	736	9	AI668782	BI711712 wcl4b12.x
c 20	563	39.8	624	14	BQ582098	AI668782 wcl4b12.x
c 21	557	39.3	559	13	BM141822	BQ582098 ill12906.y
c 22	548	38.7	550	13	BM141736	BM141822 if26h04.x
c 23	547	38.6	766	14	BM679998	BM141736 if25d10.x
c 24	545	38.5	793	12	BG436916	BM679998 UI-E-E01-
c 25	543	38.3	746	13	BI752623	BG436916 602488424
c 26	539	38.1	938	14	BQ073413	BI752623 603028471
c 27	534	37.7	764	12	BG762506	BQ073413 AGENCOURT
c 28	524	37.0	550	9	AI017543	BG762506 602733954
c 29	510	36.0	710	9	AI797380	AI017543 ou35c05.x
c 30	510	36.0	781	12	BG532265	AI797380 we87d12.x
c 31	510	36.0	928	13	BI822844	BG532265 602561355
c 32	505	35.7	537	13	BM142010	BI822844 603040166
c 33	503	35.5	552	13	BM142111	BM142010 if25d10.y
c 34	487	34.4	560	10	AW182514	BM142111 if26h04.y
c 35	468	33.1	576	10	AW956706	AW182514 xj81b03.x
c 36	468	33.1	657	13	BI916460	AW956706 EST368776
c 37	461	32.6	461	9	AA907376	BI916460 603182527
c 38	459	32.4	946	13	BI757823	AA907376 oj70h11.s
c 39	457	32.3	790	12	BG703176	BI757823 603030404
c 40	452	31.9	813	13	BI755493	BG703176 602686082
c 41	451	31.9	633	14	BQ002524	BI755493 603027349
c 42	448	31.6	791	12	BG776560	BQ002524 UI-H-E11-
c 43	446	31.5	579	9	AI559188	BG776560 602663693
c 44	444	31.4	749	10	AW073366	AI559188 tq42905.x
c 45	434	30.6	482	9	AI085503	AW073366 xa39c01.x
						AI085503 qf20c07.x

#### ALIGNMENTS

RESULT 1  
AL529152  
LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL529152 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5  
prime, mRNA sequence.  
ACCESSION AL529152  
VERSION AL529152.1 GI:12792645  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 844)  
AUTHORS Li W.B., Gruber.C., Jessee.J. and Polayes.D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
Location/Qualifiers  
1..844

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

```
BASE COUNT      182 a   238 c   235 g   189 t
ORIGIN
Query Match      59.6%; Score 844; DB 9; Length 844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 300 AAACATGTGCTGAGTGGTTCCAGAGGCTGCCAGACTGGGTGCTGCTTTCCTG 359
Db 1 AAACATGTGCTGAGTGGTTCCAGAGGCTGCCAGACTGGGTGCTGCTTTCCTG 60
QY 360 CTGAGGCAATTTGACTTCATTCCAGGGACCTTCAGAGACGCTACACTGCTGAACCA 419
Db 61 CTGAGGCAATTTGACTTCATTCCAGGGACCTTCAGAGACGCTACACTGCTGAACCA 120
QY 420 CTGGGTGGGAACATTTTGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTG 479
Db 121 CTGGGTGGGAACATTTTGAAGAATACACCCAGCTTGCCAGGGAATGTGGACTCTGGCTG 180
QY 480 TCCCTTGGGTGGTTCCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAAATCTACAAT 539
Db 181 TCCCTTGGGTGGTTCCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAAATCTACAAT 240
QY 540 TGTCACTGCTCTGAACAGCAAGAGGCGAGTAGTGGCCACTTACAGGAACACATCTG 599
Db 241 TGTCACTGCTCTGAACAGCAAGAGGCGAGTAGTGGCCACTTACAGGAACACATCTG 300
QY 600 TGTGACGTAGAGATCCAGGGCAGGGGCTATGTGTGAAGCAACTCTACCATGCTGGG 659
Db 301 TGTGACGTAGAGATCCAGGGCAGGGGCTATGTGTGAAGCAACTCTACCATGCTGGG 360
QY 660 CCGAGTCTTGAGTCACTGTCTAGCAGACACAGCAGGAGAGATTGCTAGCTCTGCTAT 719
Db 361 CCGAGTCTTGAGTCACTGTCTAGCAGACACAGCAGGAGAGATTGCTAGCTCTGCTAT 420
QY 720 GACATGCGGTTCCCTGAACCTCTCTGGCATGGCTCAAGCTGGAGCAGAGATACTTACC 779
Db 421 GACATGCGGTTCCCTGAACCTCTCTGGCATTTGGCTCAAGCTGGAGCAGAGATACTTACC 480
QY 780 TATCCTTCAGCTTTTGGATCCATTACAGGCCCGAGCCCACTGGGAGGTGTGTCGGGCC 839
Db 481 TATCCTTCAGCTTTTGGATCCATTACAGGCCCGAGCCCACTGGGAGGTGTGTCGGGCC 540
QY 840 CCGTGTATCGAAACCCAGTGTATGTAGTGGCAGCAGCAGAGTGTGGACGCCACCATGAG 899
Db 541 CCGTGTATCGAAACCCAGTGTATGTAGTGGCAGCAGCAGAGTGTGGACGCCACCATGAG 600
QY 900 AAGAGGCAAGTTATGGCCACAGCATGTGTAGACCCCTGGGGAACAGTGTGGCCCGC 959
Db 601 AAGAGGCAAGTTATGGCCACAGCATGTGTAGACCCCTGGGGAACAGTGTGGCCCGC 660
QY 960 TGCTCTGAGGGCCAGGCGCTCTGCGCTTGGCCGAATAGACCTCAACTATCTGCGACAGTTG 1019
Db 661 TGCTCTGAGGGCCAGGCGCTCTGCGCTTGGCCGAATAGACCTCAACTATCTGCGACAGTTG 720
QY 1020 CCGCAGACACTCGCTGTGTTCAGCAGCCGAGGCGTACCTCTATGGCAATGTGGGTCAAC 1079
Db 721 CCGCAGACACTCGCTGTGTTCAGCAGCCGAGGCGTACCTCTATGGCAATGTGGGTCAAC 780
QY 1080 CCACGTCTTAGACTTCTGACTTCTGTAGTTTAGACCTTGGCCCTCCACCCACCCCTTGC 1139
Db 781 CCACGTCTTAGACTTCTGACTTCTGTAGTTTAGACCTTGGCCCTCCACCCACCCCTTGC 840
```

```
QY 1140 CACT 1143
Db 841 CACT 844

RESULT 2
AL520768
LOCUS      AL520768 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB002YD05 5
DEFINITION prime, mRNA sequence.
ACCESSION  AL520768
VERSION    AL520768.1 GI:12784261
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 890)
AUTHORS   Li.W.B., Gruber.C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES             Location/Qualifiers
     source            1..890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DB002YD05"
                     /clone_lib="LTI_NFL004_NBC2"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
     note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT    191 a   251 c   241 g   206 t   1 others
ORIGIN
Query Match      57.4%; Score 813; DB 9; Length 890;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 TCATCACCAGCCCTCCACAGATTCCTGCTCCCTCTCTGTCCTGGACTCCGGATACCTC 177
Db 70 TCATCACCAGCCCTCCACAGATTCCTGCTCCCTCTCTGTCCTGGACTCCGGATACCTC 129
QY 178 AACCTCTAGTACTTGTGCTCAGCCCGAGCCAGCCATGGCTATCTCCTCTTCTCCTCT 237
Db 130 AACCTCTAGTACTTGTGCTCAGCCCGAGCCCGAGCCATGGCTATCTCCTCTTCTCCTCT 189
QY 238 GCGAACTCCCTCTGGTGGCTGTGTGCCAGGTAAACATCGACGCCAGCAAGCAAGAACT 297
Db 190 GCGAACTCCCTCTGGTGGCTGTGTGCCAGGTAAACATCGACGCCAGCAAGCAAGAACT 249
QY 298 TTAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTTCC 357
Db 250 TTAAACATGTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTTCC 309
QY 358 TGCTGTAGGCAATTTGACTTTCATTCACGGGACCTGTCAGAGACGCTACACCTGTCTGAAC 417
Db 310 TGCTGTAGGCAATTTGACTTTCATTCACGGGACCTGTCAGAGACGCTACACCTGTCTGAAC 369
QY 418 CACTGGGTGGGAACTTTTGGAGAAATACCCAGCTTGCAGGGAATGTGACTCTTGGC 477
Db 418 CACTGGGTGGGAACTTTTGGAGAAATACCCAGCTTGCAGGGAATGTGACTCTTGGC 477
```



```
prime, mRNA sequence.
ACCESSION AL522373
VERSION AL522373.1 GI:12785866
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 960)
JOURNAL Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
COMMENT Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB08YM02"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 225 a 259 c 255 g 217 t 4 others
ORIGIN
Query Match 48.8%; Score 691; DB 9; Length 960;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 911; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 438 GAAGATACACCCAGCTGGCCAGGAATGTGGACTCTGCTGCTGCTGGGTGGTTTCCAT 497
Db 922 GAAGATACACCCAGCTGGCCAGGAATGTGGACTCTGCTGCTGCTGGGTGGTTTCCAT 863
QY 498 GAGCGTGGCCAGACTGGGAGCAGACTCAGAAAATCTACAATTGTACGCTGCTGAAC 557
Db 862 GAGCGTGGCCAGACTGGGAGCAGACTCAGAAAATCTACAATTGTACGCTGCTGAAC 803
QY 558 AGCAAGGGGCGAGTAGTGCCACTTACAGGAAGACACATCTGTGTGAGCTAGAGATTCCA 617
Db 802 AGCAAGGGGCGAGTAGTGCCACTTACAGGAAGACACATCTGTGTGAGCTAGAGATTCCA 743
QY 618 GGGCAGGGGCGCTATGTGTGAAGCAACTCTACCATGCTGGGCCAGCTTGTGAGTCACCT 677
Db 742 GGGCAGGGGCGCTATGTGTGAAGCAACTCTACCATGCTGGGCCAGCTTGTGAGTCACCT 683
QY 678 GTCAGCACACACAGGAGCAAGATTGGTGTAGCTGTCTGTATGACATGGGTTCCCTGAA 737
Db 682 GTCAGCACACACAGGCAAGATTGGTGTAGCTGTCTGTATGACATGC-GTTCCCTGAA 624
QY 738 CTCTCTGTGGCATTGGCTCAAGCTGAGCAGAGATACCTATCTCTTCAGCTTTTGA 797
Db 623 CTCTCTGTGGCATTGGCTCAAGCTGAGCAGAGATACCTATCTCTTCAGCTTTTGA 564
QY 798 TCCATTACAGGCCCCAGCCACTGGGAGGTGTGTGCTGGGCCCGCTGCTATCGAAACCCAG 857
Db 563 TCCATTACAGGCCCCAGCCACTGGGAGGTGTGTGCTGGGCCCGCTGCTATCGAAACCCAG 504
QY 858 TGCTATGTAGTGGCAGCAGCAGCTGTGGACCCACCATGAGAACAGCAAGTTATGCC 917
Db 503 TGCTATGTAGTGGCAGCAGCAGCTGTGGACCCACCATGAGAACAGCAAGTTATGCC 444
```



BASE COUNT	185 a	209 c	200 g	186 t	2 others
ORIGIN					
Query Match	47.8%; Score 677; DB 14; Length 782;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches	677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	696 AAGATTGCTAGCTGCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGCGATTGGCT 755				
Db	1 AAGATTGCTAGCTGCTGCTATGACATGCGGTTCCCTGAACTCTCTCTGCGATTGGCT 60				
Qy	756 CAAGCTGGAGCAGATACCTTACCTTACCTTCAGCTTTTGGATCCATTACAGGCCCGCC 815				
Db	61 CAAGCTGGAGCAGATACCTTACCTTACCTTCAGCTTTTGGATCCATTACAGGCCCGCC 120				
Qy	816 CACTGGGAGGTGCTGCTGGGCGCGTGTATCGAAACCCAGTGTATGTAGTGGCAGCA 875				
Db	121 CACTGGGAGGTGCTGCTGGGCGCGTGTATCGAAACCCAGTGTATGTAGTGGCAGCA 180				
Qy	876 GCACAGTGTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGGTGTAGAC 935				
Db	181 GCACAGTGTGGAGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGGTGTAGAC 240				
Qy	936 CCTGGGGAACAGTGTGGCGCGTCTGTGAGGGCCAGGCCCTCTGCCCTGCCCGAATA 995				
Db	241 CCTGGGGAACAGTGTGGCGCGTCTGTGAGGGCCAGGCCCTCTGCCCTGCCCGAATA 300				
Qy	996 GACCTCAACTATCTCGACAGTTGGCGCGACACCTGCTGTTCAGCACCGCAGGCCT 1055				
Db	301 GACCTCAACTATCTCGACAGTTGGCGCGACACCTGCTGTTCAGCACCGCAGGCCT 360				
Qy	1056 GACCTCTATGGCAATCTGGGTACCCACTGTCTTAAGACTTTGACTTCTGTGAGTTAGAC 1115				
Db	361 GACCTCTATGGCAATCTGGGTACCCACTGTCTTAAGACTTTGACTTCTGTGAGTTAGAC 420				
Qy	1116 CTGCCCTCCACCCCGCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGGCAGG 1175				
Db	421 CTGCCCTCCACCCCGCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGGCAGG 480				
Qy	1176 ATCCAGGCACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGG 1235				
Db	481 ATCCAGGCACAGCTCCCTCACTTGGAGAACCTTGACTCTCTTGATGGAACACAGATGGG 540				
Qy	1236 CTGCTTGGGAAAGAACTTTACCTGAGCTTCACCTGAGGTGAGCTGAGCTTTTCAGAAA 1295				
Db	541 CTGCTTGGGAAAGAACTTTACCTGAGCTTCACCTGAGGTGAGCTGAGCTTTTCAGAAA 600				
Qy	1296 GGTGAATTTTATAGTCATTTTATTTATTCATGAAACTGAAGTCTGCTGAGGCTGA 1355				
Db	601 GGTGAATTTTATAGTCATTTTATTTATTCATGAAACTGAAGTCTGCTGAGGCTGA 660				
Qy	1356 GCAGCACTGGCATTGAA 1372				
Db	661 GCAGCACTGGCATTGAA 677				
RESULT 6					
BQ059007					
LOCUS					
DEFINITION	BQ059007 1048 bp mRNA linear EST 29-MAR-2002				
VERSION	AGENCOURT_6808192 NIH_MGC_99 Homo sapiens cdna clone IMAGE:5814981				
KEYWORDS	5', mRNA sequence.				
SOURCE	BQ059007.1 GI:19818347				
ORGANISM	human.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 1048)				
JOURNAL	NIH-MGC <a href="http://mgi.nci.nih.gov/">http://mgi.nci.nih.gov/</a> .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				

BASE COUNT	216 a	304 c	289 g	239 t
ORIGIN				
Query Match	47.7%; Score 675; DB 14; Length 1048;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches	675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	110 GCTGGGCTTCATCACCAGGCTCCTCACAGATTCTGTCCCTCTGTGCTCGGACTCGG 169			
Db	65 GCTGGGCTTCATCACCAGGCTCCTCACAGATTCTGTCCCTCTGTGCTCGGACTCGG 124			
Qy	170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCAGGCGATGGCTATCTCCTC 229			
Db	125 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCAGGCGATGGCTATCTCCTC 184			
Qy	230 TTCTCTCTCGGAACATGCCCTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 289			
Db	185 TTCTCTCTCGGAACATGCCCTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 244			
Qy	290 ACAGAACTTTAAACATGTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 349			
Db	245 ACAGAACTTTAAACATGTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 304			
Qy	350 GGCTTTCTCTGCTGAGGCTTTGACTTTCATTCGACGGGACCTTCGAGAGACCTACACCT 409			
Db	305 GGCTTTCTCTGCTGAGGCTTTGACTTTCATTCGACGGGACCTTCGAGAGACCTACACCT 364			
Qy	410 GTCTGAACACTGGTGGGAAACTTTTGGAAAGATACACCCAGCTTGGCAGGAATGTGG 469			
Db	365 GTCTGAACACTGGTGGGAAACTTTTGGAAAGATACACCCAGCTTGGCAGGAATGTGG 424			
Qy	470 ACTCTGGCTGCTTGGGTTTCCATGAGGCTGGCCAGACTGGGAGCAGACTCAGAA 529			
Db	425 ACTCTGGCTGCTTGGGTTTCCATGAGGCTGGCCAGACTGGGAGCAGACTCAGAA 484			
Qy	530 AATCTACAATTTGCTACGTGCTGCTGAACAGCAAAAGGGCAGTAGTGGCCACTTACAGGAA 589			
Db	485 AATCTACAATTTGCTACGTGCTGCTGAACAGCAAAAGGGCAGTAGTGGCCACTTACAGGAA 544			
Qy	590 GACACATCTGTGTACGTAGAGATTCCAGGGCAGGGGCTTATGTGTGAAGCAACTTAC 649			
Db	545 GACACATCTGTGTACGTAGAGATTCCAGGGCAGGGGCTTATGTGTGAAGCAACTTAC 604			
Qy	650 CATGCTGGGCGGCTCTTGTAGTCACTGTGTAGCAGCACACAGCAGGCGAGATTGGTCTAGC 709			
Db	605 CATGCTGGGCGGCTCTTGTAGTCACTGTGTAGCAGCACACAGCAGGCGAGATTGGTCTAGC 664			

Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LCM2069 row: h column: 22  
High quality sequence stop: 725.

FEATURES  
source  
1..1048  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5814981"  
/clone\_lib="NIH\_MGC\_99"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

```
QY 710 TGTCTGCTATGACATGCGGTTCCCTCAACTCTCTCTGCGATTG8CTCAAGCTGGAGCAGA 769
|||||
Db 665 TGTCTGCTATGACATGCGGTTCCCTCAACTCTCTCTGCGATTG8CTCAAGCTGGAGCAGA 724
|||||
QY 770 GATACTTACCTATCC 784
|||||
Db 725 GATACTTACCTATCC 739
|||||

RESULT 7
BQ581760/c
LOCUS BQ581760 677 bp mRNA linear EST 20-JUN-2002
DEFINITION i112g06.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029771
3' similar to TR:076091 076091 NITRILASE HOMOLOG 1. ;, mRNA
sequence.
ACCESSION BQ581760
VERSION BQ581760.1 GI:21494649
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 677)
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@iohph.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 483.
FEATURES
Location/Qualifiers
1..677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6029771"
/clone_lib="Human insulinoma"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site:2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University pancreas EST project library."
BASE COUNT 158 a 170 c 179 g 170 t
ORIGIN
Query Match 46.6%; Score 660; DB 14; Length 677;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 732 CCTGAACCTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCT 791
```

```
Db 677 CCTGAACCTCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTACCTATCCTTCAGCT 618
|||||
QY 792 TTTGGATCCATTACAGGCCAGCCAGCCACTGGAGGTGTTGCTGGCGGCCGCTGCTATCCAA 851
|||||
Db 617 TTTGGATCCATTACAGGCCAGCCAGCCACTGGAGGTGTTGCTGGCGGCCGCTGCTATCCAA 558
|||||
QY 852 ACCCAGTCTATGTAGTGGCAGCAGCAGTGTGGACGCCACCACCATGAGAAGAGAGCAAGT 911
|||||
Db 557 ACCCAGTCTATGTAGTGGCAGCAGCAGTGTGGACGCCACCACCATGAGAAGAGAGCAAGT 498
|||||
QY 912 TATGCCACACATGGTGGTATAGACCCCTGGGGAACAGTGGTGGGCCCTGCTCTGTAGGGG 971
|||||
Db 497 TATGCCACACATGGTGGTATAGACCCCTGGGGAACAGTGGTGGGCCCTGCTCTGTAGGGG 438
|||||
QY 972 CCAGGCTCTGCCCTGCCGAATAGACCTCAACTATCTGCCGACAGTTGGCGGCACACCTG 1031
|||||
Db 437 CCAGGCTCTGCCCTGCCGAATAGACCTCAACTATCTGCCGACAGTTGGCGGCACACCTG 378
|||||
QY 1032 CTGTGTTCACAGCAGCCGCTGACCTCTATGGCAATCTGGGTACCCACTGTCTTAA 1091
|||||
Db 377 CTGTGTTCACAGCAGCCGCTGACCTCTATGGCAATCTGGGTACCCACTGTCTTAA 318
|||||
QY 1092 GACTTGACTTCTGTAGTATAGACCTGCCCTCCACCCACCCCTGCCACTATGAGCTA 1151
|||||
Db 317 GACTTGACTTCTGTAGTATAGACCTGCCCTCCACCCACCCCTGCCACTATGAGCTA 258
|||||
QY 1152 GTGCTCATGTACATGGAGGAGGAGTCCAGCCAGCTGCCCTCCTCCTTGGAGAGACCTTGA 1211
|||||
Db 257 GTGCTCATGTACATGGAGGAGGAGTCCAGCCAGCTGCCCTCCTCCTTGGAGAGACCTTGA 198
|||||
QY 1212 CTCCTTTGATGAACACAGATGGGCTGCTGGGAAAGAAACTTTTCACCTGAGCTTCACCT 1271
|||||
Db 197 CTCCTTTGATGAACACAGATGGGCTGCTGGGAAAGAAACTTTTCACCTGAGCTTCACCT 138
|||||
QY 1272 GAGTTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCAATGTTTATTCATGGA 1331
|||||
Db 137 GAGTTCAGACTGCAGTTTCAGAAAGGTGGAATTTTATATAGTCAATGTTTATTCATGGA 78
|||||
QY 1332 AACTGAAGTCTCTGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAATATAATATATAAG 1391
|||||
Db 77 AACTGAAGTCTCTGCTGAGGCTGAGCAGCAGCTGGCATTGAAAAATATAATATATAAG 18
|||||

RESULT 8
BQ666784/c
LOCUS BQ666784 659 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-CL1-aez-a-24-0-UI.s1 UI-E-CL1 Homo sapiens cDNA clone
UI-E-CL1-aez-a-24-0-UI 3', mRNA sequence.
ACCESSION BQ666784
VERSION BQ666784.1 GI:18974615
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 659)
Bonald,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
```

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
The following repetitive elements were found in this cDNA  
sequence: 1-35, >AT-rich#Low\_complexity  
Seq primer: M13 Forward  
POLYA=yes.

FEATURES

source

Location/Qualifiers  
1..659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-CL1-aez-a-24-0-UI"  
/clone\_lib="UI-E-CL1"  
/tissue\_type="human retina"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: pT7n3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-CL1 is a normalized cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7n3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
TAG\_LIB=UI-E-CL1  
TAG\_TISSUE=human retina  
TAG\_SEQ=CCGCG"

BASE COUNT 153 a 168 c 173 g 165 t  
ORIGIN

Query Match 45.6%; Score 645; DB 13; Length 659;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 747 GCATTGGCTCAAGCTGGAGCAGATACCTACCTATCTTCAGCTTTGGATCCATTACA 806  
|||||  
Db 659 GCATTGGCTCAAGCTGGAGCAGATACCTACCTATCTTCAGCTTTGGATCCATTACA 600  
QY 807 GGCCAGCCCACTGGAGGTGTGTGCGGGCCCGTGTATCGCAACCCAGTGCATGTA 866  
|||||  
Db 599 GGCCAGCCCACTGGAGGTGTGTGCGGGCCCGTGTATCGCAACCCAGTGCATGTA 540  
QY 867 GTGGCAGCAGCAGTGTGGCGCCACCATGAGAGAGAGCAAGTTATGCCACAGCATG 926  
|||||  
Db 539 GTGGCAGCAGCAGTGTGGCGCCACCATGAGAGAGAGCAAGTTATGCCACAGCATG 480  
QY 927 GTGGTAGACCCCTGGGGAACAGTGGGCCCTGCTCTGAGGGCCAGGCCTTCGCTT 986  
|||||  
Db 479 GTGGTAGACCCCTGGGGAACAGTGGGCCCTGCTCTGAGGGCCAGGCCTTCGCTT 420  
QY 987 GCCGAATAGACCTCAACTATCTGCGACAGTTGCGCGGACACTGCTGTTCAGACAC 1046  
|||||  
Db 419 GCCGAATAGACCTCAACTATCTGCGACAGTTGCGCGGACACTGCTGTTCAGACAC 360  
QY 1047 CGAGGCCCTGACTCTATGGCAATCTGGGTACCCCACTGCTCTTAAGACTTGACTTCTGTG 1106  
|||||  
Db 359 CGAGGCCCTGACTCTATGGCAATCTGGGTACCCCACTGCTCTTAAGACTTGACTTCTGTG 300  
QY 1107 AGTTTAGACCTGCCCTCCACCCCACTGCCACTATGAGCTAGTGTCTATGACTTT 1166  
|||||  
Db 299 AGTTTAGACCTGCCCTCCACCCCACTGCCACTATGAGCTAGTGTCTATGACTTT 240  
QY 1167 GGAGGCAGGATCCAGGCACAGCTCCCTCCTCACTTGGGAACCTTGACTCTCTTGATGGAAC 1226  
|||||  
Db 239 GGAGGCAGGATCCAGGCACAGCTCCCTCCTCACTTGGGAACCTTGACTCTCTTGATGGAAC 180  
QY 1227 ACAGATGGGCTGCTTGGGAAGAAAAGCTTTACCTGAGCTTACCTGAGCTCAGACTGCAG 1286

Db 179 ACAGATGGGCTGCTTGGGAAGAAAAGCTTTACCTGAGCTTCACCTGAGCTCAGACTGCAG 120  
|||||  
QY 1287 TTTCAGAAAGGTGGAAATTTATATAGTCACTTCTTATTTTCATGGAACTGAAGTTCTGCT 1346  
|||||  
Db 119 TTTCAGAAAGGTGGAAATTTATATAGTCACTTCTTATTTTCATGGAACTGAAGTTCTGCT 60  
QY 1347 GAGGCTCAGCAGCAGCTGGCATTGAAAAATAATAATCATAAG 1391  
|||||  
Db 59 GAGGCTCAGCAGCAGCTGGCATTGAAAAATAATAATCATAAG 15  
RESULT 9  
BI711300/c  
LOCUS BI711300 625 bp mRNA linear EST 11-MAR-2002  
DEFINITION id96fi12 x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5023654  
3' similar to TR:076091 076091 NITRILASE HOMOLOG 1.; mRNA  
sequence.  
ACCESSION BI711300  
VERSION BI711300  
KEYWORDS EST. GI:15686995  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 625)  
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A.,  
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas  
M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,  
Jackson,Y. and Bowers,Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@bioharp.harvard.edu  
Library was constructed by Dr. J. J. Ferrer In vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
University Genome Sequencing Center For information on obtaining a  
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 481.  
Location/Qualifiers  
1..625  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5023654"  
/clone\_lib="Human insulinoma"  
/tissue\_type="insulinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:  
XhoI; Site\_2: EcoRI; Constructed with lambda ZAPII system  
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue following the Washington  
University protocol  
(http://genome.wustl.edu/est/lambda\_protocol.shtml).  
Please contact Hiroshi Inoue, MD/PhD for further  
information on this library (Metabolism Division, Permutt  
Laboratory, Washington University School of Medicine, Box  
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this  
is a Washington University Pancreas EST project library."

BASE COUNT 139 a 183 c 161 g 142 t  
ORIGIN

Query Match 44.1%; Score 625; DB 13; Length 625;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Db 255 ACTGCTCATGTGACTTGGAGGAGGATCAGGCACAGCTCCCTCACTTGGAGAACCTTG 196

QY 1211 ACTCTCTTGATGAACACAGATGGCTGCTTGGGAAAGAACTTTCACTGAGCTTCAAC 1270

Db 195 ACTCTCTTGATGAACACAGATGGCTGCTTGGGAAAGAACTTTCACTGAGCTTCAAC 136

QY 1271 TGAGGTCAAGCTGTCAGTTCAGAAAGTGGAAATTTATATAGTTCATGTTTATTTTCATGG 1330

Db 135 TGAGGTCAAGCTGTCAGTTCAGAAAGTGGAAATTTATATAGTTCATGTTTATTTTCATGG 76

QY 1331 AAACCTGAAGTTCCTGCTGAGGCTGAGCAGCACTGCATTTGAAATATAATAATCAATAA 1390

Db 75 AAACCTGAAGTTCCTGCTGAGGCTGAGCAGCACTGCATTTGAAATATAATAATCAATAA 16

QY 1391 G 1391

Db 15 G 15

RESULT 11

BM690873

LOCUS BM690873 613 bp mRNA linear EST 28-FEB-2002

DEFINITION UI-E-C11-aaz-a-06-0-UI.r1 UI-E-C11 Homo sapiens cDNA clone

ACCESSION BM690873

VERSION UI-E-C11-aaz-a-06-0-UI 5', mRNA sequence.

KEYWORDS EST.

SOURCE BM690873.1 GI:19004131

ORGANISM human.

REFERENCE 1 (bases 1 to 613)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 9704477

COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1. 613

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="UI-E-C11-aaz-a-06-0-UI"

/clone\_lib="UI-E-C11"

/tissue\_type="RPE and Choroid"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-C11 is a normalized cDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA.

This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).\*

BASE COUNT 128 a 184 c 164 g 137 t

ORIGIN

Query Match 42.8%; Score 606; DB 14; Length 613;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 645 TCTACCATGCTGGGCCAGCTCTTGAGTCACCTGTCAGACACACAGGCAAGATTTGGT 704

Db 8 TCTACCATGCTGGGCCAGCTCTTGAGTCACCTGTCAGACACACAGGCAAGATTTGGT 67

QY 705 CTAGCTGTCTGATGACATGCGTTCCCTGAACCTCTCTCTGGCATTTGGCTCAAGCTGA 764

Db 68 CTAGCTGTCTGATGACATGCGTTCCCTGAACCTCTCTCTGGCATTTGGCTCAAGCTGA 127

QY 765 GCAGAGATACCTTACCTATCCTTTCAGCTTTTGGATCCATTACAGGCCAGCCACCTGGGAG 824

Db 128 GCAGAGATACCTTACCTATCCTTTCAGCTTTTGGATCCATTACAGGCCAGCCACCTGGGAG 187

QY 825 GTGTTGCTGGGGCCCGCTGCTATCGAAACCCAGTCTCTATGAGTGGCAGCAGCAGTGT 884

Db 188 GTGTTGCTGGGGCCCGCTGCTATCGAAACCCAGTCTCTATGAGTGGCAGCAGCAGTGT 247

QY 885 GGACGCCACCATGAGAGAGAGCAAGTTATGCCACACAGCATGGTGGTAGACCCCTGGGGA 944

Db 248 GGACGCCACCATGAGAGAGAGCAAGTTATGCCACACAGCATGGTGGTAGACCCCTGGGGA 307

QY 945 ACAGTGGTGGCCCGCTGCTCTGAGGGGCGAGGCCCTCTGCCTTGCCTCGAATAGACCTCAAC 1004

Db 308 ACAGTGGTGGCCCGCTGCTCTGAGGGGCGAGGCCCTCTGCCTTGCCTCGAATAGACCTCAAC 367

QY 1005 TATCTGGACAGTGTCCGCGGACACCTGCTGTGTTCAGCACCGCAGGCTGACCTCTAT 1064

Db 368 TATCTGGACAGTGTCCGCGGACACCTGCTGTGTTCAGCACCGCAGGCTGACCTCTAT 427

QY 1065 GGCAATCTGGGTCACCCACTGCTTTAAGACTTTGACTTCTGTGAGTTTAGACCTGCCCTC 1124

Db 428 GGCAATCTGGGTCACCCACTGCTTTAAGACTTTGACTTCTGTGAGTTTAGACCTGCCCTC 487

QY 1125 CCACCCCAACCTGCCACTATGAGCTAGTGTCTATGTGACTTTGGAGGCAAGATCCAGGCA 1184

Db 488 CCACCCCAACCTGCCACTATGAGCTAGTGTCTATGTGACTTTGGAGGCAAGATCCAGGCA 547

QY 1185 CAGCTCCCTCACTTGGAGAACCTTGTGCTCTCTTATGGAACACAGATGGGCTGCTGGG 1244

Db 548 CAGCTCCCTCACTTGGAGAACCTTGTGCTCTCTTATGGAACACAGATGGGCTGCTGGG 607

QY 1245 AAAGAA 1250

Db 608 AAAGAA 613

RESULT 12

BI769604

LOCUS BI769604 846 bp mRNA linear EST 25-SEP-2001

DEFINITION 603054985F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5204469 5', mRNA sequence.

ACCESSION BI769604

VERSION BI769604.1 GI:15761182

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 846)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM1512 row: j column: 22  
High quality sequence stop: 838.  
Location/Qualifiers  
1. 846  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5204469"  
/clone\_lib="NIH\_MGC\_122"  
/lab\_host="DH10B"  
/note="organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH\_MGC Library."  
179 a 240 c 226 g 201 t  
BASE COUNT  
ORIGIN  
Query Match 42.7%; Score 604; DB 13; Length 846;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 110 GCTGGCTTCATCACAGGCTCCTCACAGATTCCTGCTCCCTCTCTGTCCTGGACTCG 169  
DB 61 GCTGGCTTCATCACAGGCTCCTCACAGATTCCTGCTCCCTCTCTGTCCTGGACTCG 120  
QY 170 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCAGAGCATGGCTATCTCCTC 229  
DB 121 GATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCAGAGCATGGCTATCTCCTC 180  
QY 230 TTCCTCTCGGAATGCCCTGGTGGCTGTGTGCAGGTAACTACATGACGCCAGACGA 289  
DB 181 TTCCTCTCGGAATGCCCTGGTGGCTGTGTGCAGGTAACTACATGACGCCAGACGA 240  
QY 290 ACAGAACTTTAAACATGTGCTGAGCTGCTCGAGAGCTGCGAGACTGGGTGCTGCT 349  
DB 241 ACAGAACTTTAAACATGTGCTGAGCTGCTCGAGAGCTGCGAGACTGGGTGCTGCT 300  
QY 350 GCTTTCTGCTGAGGCACTTTGACTTCATTCAGCGGACCTGCGAGAGCCTACACCT 409  
DB 301 GCTTTCTGCTGAGGCACTTTGACTTCATTCAGCGGACCTGCGAGAGCCTACACCT 360  
QY 410 GTCTGACCACTGGTGGGAACCTTTGGAAGATACACCCAGCTGTCAGGGAATGTGG 469  
DB 361 GTCTGACCACTGGTGGGAACCTTTGGAAGATACACCCAGCTGTCAGGGAATGTGG 420  
QY 470 ACTCTGGCTGCTCTTGGGTGGTTTCCATGAGCGTGGCCAAAGACTGGGAGCACTCAGAA 529  
DB 421 ACTCTGGCTGCTCTTGGGTGGTTTCCATGAGCGTGGCCAAAGACTGGGAGCACTCAGAA 480  
QY 530 AATCTACAATTTCACTGCTGCTGCTGAACAGCAAAAGGGCAGTAGTGGCCACTTACAGAA 589  
DB 481 AATCTACAATTTCACTGCTGCTGCTGAACAGCAAAAGGGCAGTAGTGGCCACTTACAGAA 540  
QY 590 GACATCTGTGTGACGTAGAGATTCAGGGCAGGGGCTATGTGTGAACCACTTAC 649  
DB 541 GACATCTGTGTGACGTAGAGATTCAGGGCAGGGGCTATGTGTGAACCACTTAC 600  
QY 650 CATGCTGGGGCCAGTCTTGTAGCTACCTGTCTAGCACACAGCAGGCAAGATTGGTCTAGC 709  
DB 601 CATGCTGGGGCCAGTCTTGTAGCTACCTGTCTAGCACACAGCAGGCAAGATTGGTCTAGC 660

QY 710 TGTC 713  
DB 661 TGTC 664  
RESULT 13  
BM979227/c  
LOCUS  
DEFINITION  
UI-CF-DU1-adl-n-04-0-UI.s1 750 bp mRNA linear EST 21-MAR-2002  
UI-CF-DU1-adl-n-04-0-UI 3', mRNA sequence.  
ACCESSION  
BM979227  
VERSION  
BM979227.1 GI:19599461  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (Bases 1 to 750)  
AUTHORS  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL  
Genome Res. 6 (9), 791-806 (1996)  
MEDLINE  
9704477  
COMMENT  
Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
The following repetitive elements were found in this cDNA  
sequence: 1-41, >AT-rich#Low\_complexity  
Seq primer: M13 FORWARD  
POLYA-Yes.  
FEATURES  
Location/Qualifiers  
1. 750  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-CF-DU1-adl-n-04-0-UI"  
/clone\_lib="UI-CF-DU1"  
/tissue\_type="Primary Lung Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: Lung; Vector: pT7T3-pac (Pharmacia) with a  
modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
UI-CF-DU1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Epithelial Cells The  
library was constructed according to Bonaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an EcoR I adaptor, digested with Not I, and cloned  
directionally into pT7T3-Pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (dT)18 tail. The sequence tag for this  
library is GGCTGTAGGC.  
TAG\_LIB=UI-CF-DU1  
TAG\_TISSUE=Lung Epithelial Cells Tissue nos 359-368  
TAG\_SEQ=GGCTGTAGGC"  
185 a 171 c 193 g 200 t 1 others  
BASE COUNT  
ORIGIN  
Query Match 41.6%; Score 589; DB 14; Length 750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 823 AGGTGTTGCTGGGGCCGCTGCTATCGAAACCCAGTGTATAGTGGCAGCAGCACAGT 882
Db 589 AGGTGTTGCTGGGGCCGCTGCTATCGAAACCCAGTGTATAGTGGCAGCAGCACAGT 530
QY 883 GTGGAGCCACCATGAGAAAGAGAGCAAGTTATGGCCACAGCATGTTGGTAGACCCTGGG 942
Db 529 GTGGAGCCACCATGAGAAAGAGAGCAAGTTATGGCCACAGCATGTTGGTAGACCCTGGG 470
QY 943 GAACAGTGTGGCCCGCTGCTCTGAGGGCCAGGCTCTGCTGCCGTTGCCGAATAGACCTCA 1002
Db 469 GAACAGTGTGGCCCGCTGCTCTGAGGGCCAGGCTCTGCTGCCGTTGCCGAATAGACCTCA 410
QY 1003 ACTATCTGGCAGATTGGCCGACACCTGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTT 1062
Db 409 ACTATCTGGCAGATTGGCCGACACCTGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTT 350
QY 1063 ATGGCAATCTGGGTCAACCACTGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTT 1122
Db 349 ATGGCAATCTGGGTCAACCACTGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTT 290
QY 1123 TCCCAACCCACCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGG 1182
Db 289 TCCCAACCCACCCCTGCCACTATGAGCTAGTGTCTCATGTGACTTGGAGCAGGATCCAGG 230
QY 1183 CACAGTCCCTCACTTGGAGAACCTTGACTCTCTTGTATGGAACACAGATGGCTGCTTG 1242
Db 229 CACAGTCCCTCACTTGGAGAACCTTGACTCTCTTGTATGGAACACAGATGGCTGCTTG 170
QY 1243 GGAAGAACTTTCACCTGAGCTTCACTGAGGTGAGTGTCTGCTGAGGCTGAGCAGCAC 1362
Db 169 GGAAGAACTTTCACCTGAGCTTCACTGAGGTGAGTGTCTGCTGAGGCTGAGCAGCAC 110
QY 1303 TTTTATATAGTCATGTTTATTTATGAACTGAACTGAACTGAACTGAACTGAACTGAACT 1362
Db 109 TTTTATATAGTCATGTTTATTTTATGAACTGAACTGAACTGAACTGAACTGAACT 110
QY 1363 TGGCATTTGAAATATATAATATATAATGTAAGTCAAAATGTAAGTCAAAATGTAAGT 1411
Db 49 TGGCATTTGAAATATATAATATATAATGTAAGTCAAAATGTAAGTCAAAATGTAAGT 1

RESULT 14
AL520767/c
LOCUS
DEFINITION
AL520767 LTI_NFL004_NBC2 Homo sapiens cDNA clone EST 13-FEB-2001
prime, mRNA sequence.
ACCESSION
AL520767
VERSION
AL520767.1 GI:12784260
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 992)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
CONTACT
Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB002YD05"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue.type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMWSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
```

```
cloned into the Not I and Eco RV sites of the pCMWSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 229 a 267 c 268 g 222 t 6 others
ORIGIN
Query Match 41.3%; Score 585; DB 9; Length 992;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 805; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 443 ATACACCAGCTTGGCAGGGAATGTGGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCG 502
Db 929 ATACACCAGCTTGGCAGGGAATGTGGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCG 870
QY 503 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTTGCTACGCTGCTGCTGAACAGCAA 562
Db 869 TGGCCAAGACTGGGAGCAGACTCAGAAAATCTACAATTTGCTACGCTGCTGCTGAACAGCAA 810
QY 563 AGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 622
Db 809 AGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCA 750
QY 623 GGGGCTATGTCTGAAAGCAACTCTACCATGCTGGGCCAGTCTTGGAGTACCTCTCAG 682
Db 749 GGGGCTATGTGTGAAAGCAACTCTACCATGCTGGGCCAGTCTTGGAGTACCTCTCAG 690
QY 683 CACACAGCAGCAAGATTGGTCTAGCTGTCTGTATGACATGCGGTTCCTTGAACCTCTC 742
Db 689 CACACAGCAGCAAGATTGGTCTAGCTGTCTGTATGACATGCGGTTCCTTGAACCTCTC 630
QY 743 TCTGGCATTTGGCTCAAGCTGGAGCAGAGATACCTATACCTTACGCTTTGGATCCAT 802
Db 629 TCTGGCATTTGGCTCAAGCTGGAGCAGAGATACCTATACCTTACGCTTTGGATCCAT 570
QY 803 TACAGGCCAGCCCACTGGGAGGTGTGCTGGGCCGCTGTATCGAAACCCAGTGCTA 862
Db 569 TACAGGCCAGCCCACTGGGAGGTGTGCTGGGCCGCTGTATCGAAACCCAGTGCTA 510
QY 863 TCTAGTGGC - AGCAGCAGCAGTGTGGAGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 921
Db 509 TCTAGTGGCCAGCAGCAGCAGTGTGGAGCCACCATGAGAAGAGAGCAAGTTATGGCCACA 450
QY 922 GCATGGTGTACACCCCTGGGAAACAGTGGTGGCCGCTGTCTGAGGGGCCAGGCTCT 981
Db 449 GCATGGTGTAGACCCCTGGGAAACAGTGGTGGCCGCTGTCTGAGGGGCCAGGCTCT 390
QY 982 GCCTTGGCCGAATAGACCTCAACTATCTCGACAGTTCGCCGACACCTGCCTGTGTTCC 1041
Db 389 GCCTTGGCCGAATAGACCTCAACTATCTCGACAGTTCGCCGACACCTGCCTGTGTTCC 330
QY 1042 AGCACCAGCAGGCTGACCTCTATGGCAATCTGGGTCAACCATGCTTAAAGACTTGACTT 1101
Db 329 AGCACCAGCAGGCTGACCTCTATGGCAATCTGGGTCAACCATGCTTAAAGACTTGACTT 270
QY 1102 CTGTGAGTTTACACCTGCCCCCTCCACCCACCCCTGACCTATGAGCTAGTGCATGTT 1161
Db 269 CTGTGAGTTTACACCTGCCCCCTCCACCCACCCCTGACCTATGAGCTAGTGCATGTT 210
QY 1162 GACTTGGGAGGAGATCCAGGACACAGTCCCTCCTGAGAACTTGGAACTTCTCTTGAT 1221
Db 209 GACTTGGGAGGAGATCCAGGACACAGTCCCTCCTGAGAACTTGGAACTTCTCTTGAT 1221
QY 1222 GGAACACAGATGGGTGCTTGGGAAAGA 1249
Db 149 GGAACACAGATGGGTGCTTGGGAAAGA 122
RESULT 15
AL797259/c
```



LOCUS AI797259 586 bp mRNA linear EST 18-DEC-1999  
DEFINITION web6e02.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2347994 3' similar to TR:076091 076091 NITRILASE HOMOLOG 1.  
; mRNA sequence.  
ACCESSION AI797259  
VERSION AI797259.1 GI:5362731  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 586)  
REFERENCE NC-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1500 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 468.  
Location/Qualifiers  
1..586  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2347994"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCLCGAP-GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 137 a 155 c 158 g 136 t  
ORIGIN  
Query Match 41.2%; Score 584; DB 9; Length 586;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 810 CCAGCCCACTGGGAGGTGTGCTGGGGCCGCTGCTATCGAAACCCAGTGTATGTAGTG 869  
|||||  
DB 586 CCAGCCCACTGGGAGGTGTGCTGGGGCCGCTGCTATCGAAACCCAGTGTATGTAGTG 527  
QY 870 GCAGCAGCACAGTGTGGACGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGTGTG 929  
|||||  
DB 526 GCAGCAGCACAGTGTGGACGCCACCATGAGAGAGAGCAAGTTATGGCCACAGCATGTGTG 467  
QY 930 GTAGACCCCTGGGAAACAGTGTGGCCCGCTGCTGAGGGGCCAGGCTCTGCCCTGCC 989  
|||||  
DB 466 GTAGACCCCTGGGAAACAGTGTGGCCCGCTGCTGAGGGGCCAGGCTCTGCCCTGCC 407  
QY 990 CGAATAGACCTCAACTATCTGCGACAGTTGGCCCGCACACCTGCCCTGTGTTCAGCACCCGC 1049  
|||||  
DB 406 CGAATAGACCTCAACTATCTGCGACAGTTGGCCCGCACACCTGCCCTGTGTTCAGCACCCGC 347  
QY 1050 AGGCTGACCTCTATGGCAATCTGGGTCAACCACCTGTCTTAAGACTTGACTTCTGTGAGT 1109  
|||||  
DB 346 AGGCTGACCTCTATGGCAATCTGGGTCAACCACCTGTCTTAAGACTTGACTTCTGTGAGT 287  
QY 1110 TTAGACCTGCCCTCCACCCCTCCACCCCTGAGCTAGTGTCTATGTGACTTGGA 1169  
|||||  
DB 286 TTAGACCTGCCCTCCACCCCTCCACCCCTGAGCTAGTGTCTATGTGACTTGGA 227  
..

QY 1170 GGAGGATCCAGGCACAGCTCCCTCACTTGGAGAACCTTCACTCTCTTGTATGGAACACA 1229  
|||||  
DB 226 GGAGGATCCAGGCACAGCTCCCTCACTTGGAGAACCTTCACTCTCTTGTATGGAACACA 167  
QY 1230 GATGGCTGCTTGGGAAAGAACTTTTACACCTGAGCTTCCACCTGAGGTGAGGTGAGTTC 1289  
|||||  
DB 166 GATGGCTGCTTGGGAAAGAACTTTTACACCTGAGCTTCCACCTGAGGTGAGGTGAGTTC 107  
QY 1290 CAGAAAGTGGAAATTTTATATAGTCATTTGTTTATTTTCATGGAACCTGAAGTTCTGCTGAG 1349  
|||||  
DB 106 CAGAAAGTGGAAATTTTATATAGTCATTTGTTTATTTTCATGGAACCTGAAGTTCTGCTGAG 47  
QY 1350 GGCTGAGCAGCAGCTGGGCATTGAAAAATAATAATAATCAATAAAGTC 1393  
|||||  
DB 46 GGCTGAGCAGCAGCTGGGCATTGAAAAATAATAATAATCAATAAAGTC 3

Search completed: November 30, 2002, 14:00:56  
Job time : 1870 secs



OM of: US-09-357-675C-21 to: GenEmbl:\* out\_format : pfs

Date: Apr 29, 2002 11:05 AM

About: Results were produced by the GenCore software, version 4.5,

Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09357675/runat.29042002.091757.3179/app\_query.fasta.1.1208  
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM\_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675\_CGNI\_1\_10065 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-357-675C-21

Query length: 327

Database: GenEmbl.\*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 3374.980000

score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb.pr:AF069987	+ 1748.00	2415.67	3.4e-126	1385	! AF069987 Homo sapiens nitrilas
gb.ro:AF069988	+ 1476.00	2049.81	8.2e-106	1338	! AF069988 Mus musculus nitrilas
gb.ro:BC021634	+ 1476.00	2038.53	3.5e-104	1365	! BC021634 Mus musculus, nitrila
gb.pr:AF069984	+ 1468.50	2018.45	4.6e-104	4079	! AF069984 Homo sapiens nitrilas
gb.htg:AL354714.4	+ 1468.50	1989.30	1.9e-102	110000	! Continuation (5 of 7) of AL3
gb.htg:AL354714.5	+ 1468.50	1989.30	1.9e-102	110000	! Continuation (6 of 7) of AL3
gb.htg:AL590651	+ 1468.50	1985.56	3.1e-102	167863	! AL590651 Homo sapiens chromo
gb.pr:AL591806	+ 1468.50	1983.97	3.8e-102	200822	! AL591806 Human DNA sequence
gb.htg:AC084821	+ 1105.50	1480.02	4.5e-74	211771	! AC084821 Mus musculus chromo
gb.ro:AF069985	+ 1102.50	1509.97	9.6e-76	4481	! AF069985 Mus musculus nitrilas
gb.htg:AC087229	+ 1083.00	1448.75	2.5e-72	213203	! AC087229 Mus musculus chromo
gb.pr:AC105589	+ 1059.00	1417.03	1.4e-70	178418	! AC105589 Rattus norvegicus
gb.ov:AF284575	+ 968.50	1335.66	4.9e-66	1214	! AF284575 Xenopus laevis Nit pr
gb.in:AF069989	+ 710.50	975.82	5.4e-46	1521	! AF069989 Drosophila melanogast
gb.htg:AC017755	+ 710.50	936.34	8.6e-44	131853	! AC017755 Drosophila melanoga
gb.in:AC093121	+ 710.50	934.18	1.1e-43	168250	! AC093121 Drosophila melanoga
gb.in:AC093501	+ 710.50	933.91	1.2e-43	173494	! AC093501 Drosophila melanoga
gb.htg:AC006169	+ 710.50	933.87	1.2e-43	174279	! AC006169 Drosophila melanoga
gb.in:AC105293	+ 710.50	933.09	1.3e-43	190351	! AC105293 Drosophila melanoga
gb.in:AE003467	+ 710.50	929.10	2.2e-43	298640	! AE003467 Drosophila melanoga
gb.in:AF069986	+ 702.00	964.86	2.2e-45	1385	! AF069986 Caenorhabditis elega
gb.pl:AF372904	+ 648.50	892.13	2.5e-41	1172	! AF372904 Arabidopsis thaliana
gb.htg:AF010013	+ 632.50	840.14	2.0e-38	34013	! AF010013 Drosophila melanog
gb.htg:SCXJL126W	+ 519.00	715.54	1.7e-31	15124	! AE000934 Methanobacterium the
gb.pl:SCXJL126W	+ 519.00	715.54	1.7e-31	15124	! AE000934 Methanobacterium the
gb.pl:AF284571	+ 519.00	709.80	3.6e-31	1307	! AF284571 Saccharomyces cerevis
gb.pl:SPRC651	+ 502.00	661.73	1.7e-28	23550	! AL035570 S.pombe chromosome I
gb.in:AE0004861	+ 500.50	667.48	8.1e-29	10466	! AE0004861 Pseudomonas aerugin
gb.pr:AF260334	+ 493.00	677.68	2.2e-29	1019	! AF260334 Homo sapiens CUA002 m
gb.pl:SCGCD14	+ 493.00	668.77	6.9e-29	2792	! 254149 S.cerevisiae GCD14 gene
gb.ba:SYCSLRD	+ 490.00	632.03	7.7e-27	110908	! D64002 Synecocystis sp. PCC
gb.pr:BC020620	+ 487.50	670.54	5.5e-29	965	! BC020620 Homo sapiens, Nit prot
gb.pr:AF284574	+ 485.50	667.79	7.8e-29	962	! AF284574 Homo sapiens Nit prote
gb.ro:AF284573	+ 479.00	658.94	2.4e-28	1292	! AF284573 Mus musculus Nit prot
gb.ba:AB017194	+ 479.00	645.32	1.4e-27	4762	! AB017194 Plectonoma boryanum C
gb.ro:BC020153	+ 479.00	657.89	2.8e-28	1063	! BC020153 Mus musculus, RIKEN c
gb.ro:AP003587	+ 477.50	604.69	2.6e-25	343550	! AP003587 Nostoc sp. PCC 7120
gb.ba:AF005512	+ 476.50	644.73	1.5e-27	3180	! AF005512 Nostoc commune CRNA-1
gb.sts:G72919	+ 471.00	648.81	8.9e-28	847	! G72919 MARC 2849-2850:99193517
gb.ba:AL646071	+ 468.00	596.16	7.6e-25	203050	! AL646071 Ralstonia solanacea

gb.ba:SME591791 - 444.50 558.98 9.0e-23 340900 ! AL591791 Sinorhizobium me  
gb.ba:AJ414158 + 440.50 556.73 1.2e-22 235050 ! AJ414158 Yersinia pestis  
gb.ba:AE004129 + 436.50 578.36 7.5e-24 10881 ! AE004129 Vibrio cholerae c  
gb.ba:AE005759 - 433.50 572.80 1.5e-23 12748 ! AE005759 Caulobacter cresc  
gb.pl:SPAC26A3 + 430.50 558.94 9.0e-23 38149 ! Z69240 S.pombe chromosome

seq\_name: gb\_pr:AF069987

seq\_documentation block:

LOCUS AF069987 1385 bp mRNA linear PRI 23-JUL-1998  
DEFINITION Homo sapiens nitrilase 1 (NIT1) mRNA, complete cds.  
ACCESSION AF069987  
VERSION AF069987.1 GI:3228665  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1385)  
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Nitrilase and fhit homologs are encoded as fusion proteins in  
drosophila melanogaster and caenorhabditis elegans  
Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
MEDLINE 98337986  
REFERENCE 2 (bases 1 to 1385)  
AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Direct Submission  
Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

FEATURES

Source  
1..1385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="lq21-q22"  
1..1385  
/gene="NIT1"  
77..1060  
/gene="NIT1"  
/note="NIT1"  
/codon\_start=1  
/product="nitrilase 1"  
/protein\_id="AAC3907.1"  
/db\_xref="GI:3228666"  
/translation="MLGFTIRPPHRLSLCPGLRIPQLSVLCAQPRPRAMASSSSC  
ELPLVAVQCVSTPDKQNFCTCAELVREAAALGACLAFLPEAFDIADPAETLHLS  
EPLGKLLLEVTQLAKREGCLMLSLGPFHERGQDWQETOKIYNCHVLLNSKGAVVATYR  
KTHLCDVIPGQGPCNSNTMPGSLSPSTPAGKIGLANCYDMRPFELSLAQA  
CAELITYPSAGSTIGTPAHWEVLLRAIETQCYVYVAAACQGRHHRKRSYGHSMVWD  
PWGTVARCSGPGCLCLARIDNLNLRRLRHLRFQHRRLPDLYNLGLHPLS"  
BASE COUNT 322 a 382 c 357 g 324 t  
ORIGIN

alignment\_scores:

Quality: 1748.00 Length: 327  
Ratio: 5.346 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-357-675C-21 x AF069987  
Align seg 1/1 to: AF069987 from: 1 to: 1385  
1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17  
77 ATGCTGGCTTCATCACGAGCTCTTCACAGATCTCTGCTCTGCTG 126  
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgp 34

```
|||||
127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGC 176
|||||
34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50
|||||
177 CCAGGACCATGGGTATCTCTCTCTCCGCAACTGCCCTGGTGGCT 226
|||||
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelLysThrCy 67
|||||
227 GTGTGGCAGGTACATCGACGCCAGACAGCAACAGAACTTTAAACATG 276
|||||
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
277 TGCTGAGCTGGTTCGAGAGAGCTCCAGACTGGGTGGCTGCTGGCTTTC 326
|||||
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
327 TGCTTGAGGATTTGACTTCAATTCGACGGAGCCCTGCAGAGAGCTACAC 376
|||||
101 LeuSerGluProLeuGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
|||||
377 CTGTCTGAACCACTGGGTGGGAACTTTTGGAGAATATACCCAGCTTGC 426
|||||
117 aArgGluCysGlyLeuThrLeuSerLeuGlyGlyPheHisGluArgGlyG 134
|||||
427 CAGGGAATGTGGACTCTGGCTGCTCTTGGGTGGTTCCTCATGAGCGTGC 476
|||||
134 lNasPTpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
|||||
477 AAGACTGGGAGCAGACTCAGAAAATCTACAAATGTCACGTGCTGCTGAAC 526
|||||
151 SerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeuCysAspVa 167
|||||
527 AGCAAGGGCAGTAGTGGCCACTTACAGGAACACACATCTGTGTGACGT 576
|||||
167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
|||||
577 AGAGATTCCAGGCAGGGGCTTATGTGTGAAGCAACTCTACCATGCCCTG 626
|||||
184 lProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
627 GGCCCAAGCTTGTAGTCACTTGTGTGAGCACTCAGCAGCAGCAAGATTG 676
|||||
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
|||||
677 GCTGTCTGTATACATGCGGTTCCCTGAACTCTCTCTGGCATTTGGCTCA 726
|||||
217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
|||||
727 AGCTGGAGCAGAGATACTTACCTATCCTTCCAGCTTTTGGATCCATTACAG 776
|||||
234 lProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
|||||
777 GCCCAGCCCACTGGGAGGTGCTGCTGGGCCCTGCTATCGAAACCCAG 826
|||||
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisGluLysArgAl 267
|||||
827 TGTATGTAGTGCAGCAGCAGCAGTGTGGACGCCACCATGAGAGAGAGC 876
|||||
267 aSerTyrGlyHisSerMetValValAspProTyrpGlyThrValValAla 284
|||||
877 AAGTTATGGCCACAGCATGTGTGTAGACCCCTGGGGAACACAGTGGGCC 926
|||||
284 rGysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
|||||
927 GCTGCTCTGAGGGGCCAGGCTCTGCTTGTCCCGGAATAGACCTCAACTAT 976
|||||
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317
|||||
977 CTGGACAGTTGGCGGACACCTTGTCTGTCTCCAGCCGAGGCGCTGA 1026
|||||
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
```

1027 CCTCTATGGCAATCTGGGTCACCCACTGTCT 1057

seq\_name: gb\_ro:AF069988

seq\_documentation\_block:

LOCUS AF069988 1338 bp mRNA linear ROD 23-JUL-1998

DEFINITION Mus musculus nitrilase 1 (Niti) mRNA, complete cds.

ACCESSION AF069988

VERSION AF069988.1 GI:3228667

KEYWORDS house mouse.

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 1338)

Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Wernuth,P., Rothman,J.H., Huebner,K.,

Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Nitrilase and fhit homologs are encoded as fusion proteins in

drosophila melanogaster and caenorhabditis elegans

Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

JOURNAL

MEDLINE

REFERENCE

AUTHORS

2 (bases 1 to 1338)

Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Wernuth,P., Rothman,J., Huebner,K.,

Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Direct Submission

Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson

Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

FEATURES

Location/Qualifiers

1..1338

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="1"

/map="1q21-q23"

1..1338

/gene="Niti"

58..1029

/gene="Niti"

/codon\_start=1

/product="nitrilase 1"

/protein\_id="AAC40185.1"

/db\_xref="GI:3228668"

/translation="MLGFTIRPPHQLCTGYRLRLTPVLCTQPRPTMSSSTSWELPL

VAVCQVSTPNKQENFKTCAELVQEARLGLACLAFLPEAFDFIARNPAETILLSEPLN

GDLEGGISQLARECGIWLISGGFHERGQDWEQNKITYNCHVLLNKGSVVASYRKTHL

CDVEIPGGQPMRESNYTKPGTLEPPVKTPAGKYGLAICIDMRFPESLSKLAQAGAEI

LTYPSAFGSVTGPAHWEYLLRRARAIESQCYVIAAAQCGRRHETRASVGHSMVDPWGT

VVARCSGPGLCGLARIDILHFLQMRQLPVPFQHRRPDLVYSLGHPLS"

BASE COUNT 347 a 335 c 330 g 326 t

ORIGIN

alignment\_scores:

Quality: 1484.00 Length: 327

Ratio: 4.834 Gaps: 2

Percent Similarity: 93.884 Percent Identity: 84.404

alignment\_block:

US-09-357-675C-21 x AF069988 ..

Align seg 1/1 to: AF069988 from: 1 to: 1338

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17

|||||

58 ATGCTGGGCTTCATCACCAGGCTCTCCACCAA.....CTCCTGTG 98

17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgp 34

|||||

99 TACCGGATACCGATTACTTCGAACCCAGTACTTTGTACTCAGCCAGGC 148

34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50

|||||

|||||

149 CCAGAACCATGTGCC...TCATCAACTTCCTGGGAGCTGCCCCCTGGTGCT 195  
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelysThrCy 67  
196 GTGTGCCAGTAAATCAATCAACACCAACCAAGAGAACTTTAAACATG 245  
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
246 TGTGTAGTTGGTTCAGAGAGCTGCCAGTGGGTGCTTGCCTGGCCTTC 295  
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
296 TGCTGAGGCAATTTGACTTTATTGCAGAAACCTGCCGAGACATTACT 345  
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluThrThrGlnLeuAl 117  
346 CTGTCCGAACCACTGAATGGGATCTTTTGGGCCAATATAGCCAGCTGC 395  
117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGly 134  
396 CAGGGAATGTGGAATCTGCTGCTCTTGGGCGGTTTCCACGAGCGTGCC 445  
134 lnAspTrpGluGlnThrGlnLysIleTyRsnCysHisValLeuLeuAsn 150  
446 AGACTGGGAGCAGAAATCAGAAATCTACAAATGTGCTATGCTTTTGAAC 495  
151 SerLysGlyAlaValValAlaThrTyRArgLysThrHisLeuCysaspVa 167  
496 AGCAAGGATCAGTAGTGCCAGTTACAGGAACACACATCTGTGCGATGT 545  
167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
546 AGAGATCCCAAGTCAGGGCGCGATGAGAGAAAGCAACTATACCAAGCCTG 595  
184 lProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
596 GAGGACCTCTTGAGCCACCTGTCTCAAGACCCGCTGGCAAGGTGGTCTA 645  
201 AlaValCysTyRAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
646 GCAATCTGTTATGACATGCGGTTCCCTGAACTTCTTGAATGGCTCA 695  
217 nAlaGlyAlaGluLeuThrTyRProSerAlaPheGlySerIleThrG 234  
696 AGCTGGGCGAGAAATCTACTTATCTTTCAGCCTTTGGATCTGTACAG 745  
234 lProAlaHisTrpGluValLeuLeuArgAlaAlaIleGluThrGln 250  
746 GTCCAGCCCACTGGGAGGTGCTGCTGGGCGCCCGCCCATTTGAATCTCAG 795  
251 CysTyRValValAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267  
796 TGCTATGTAATAGCAGCGCAGCTGTGGACGCCACCATGAACAGAGC 845  
267 aSerTyRGlyHisSerMetValValAspProTrpGlyThrValValAla 284  
846 AAGTTATGGCATAGCATGTGTGTACCCGCTGGGCGCACAGTGTGGCC 895  
284 rGysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTy 300  
896 GCTGCTCCGAGGACCAAGGCTCTGCTTGTGCAATGATGATCTCCACTT 945  
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317  
946 CTACAACAGATGCGCCAAACCTGCTGTTCAGCAGCCGACGACCTGA 995  
317 pLeuTyRGlyAsnLeuGlyHisProLeuSer 327  
996 CCTCTATGGCAGTCTGGGCTCATCCACTCTCT 1026

seq\_name: gb\_ro:BC021634

seq\_documentation\_block:

LOCUS BC021634 1365 bp mRNA linear ROD 22-JAN-2002  
DEFINITION Mus musculus, nitrilase 1, clone MGC:13825 IMAGE:4008543, mRNA,  
complete cds.  
ACCESSION BC021634  
VERSION BC021634.1 GI:18204912  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 1365)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
COMMENT Contact: MGC help desk  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
http://www.systemsbio.org  
contact: amadan@systemsbiology.org  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAK Plate: 18 Row: e Column: 20  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 6754855.  
FEATURES  
Location/Qualifiers  
1..1365  
/organism="Mus musculus"  
/db\_xref="LocusID:27045"  
/db\_xref="Taxon:10090"  
/clone="MGC:13825 IMAGE:4008543"  
/tissue\_type="Mammary tumor metastasized to lung."  
MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR  
enhancer."  
/clone\_lib="NCI\_CGAP\_Lu30"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
59..1030  
/codon\_start=1  
/product="nitrilase 1"  
/protein\_id="AAH21634.1"  
/db\_xref="GI:18204913"  
/translation="MLGFTRPQHLLCTGYRLLETPVLCTQPRPTWSSSTSWELPL  
VAVCQVSTPNKQENFKCAELVQEAARIGACLAFLPAFDIARNPAETLLLSPLN  
GDLLQYSOLARECQIWSLGGFHERGDWQNKIYNCHVLLNLSKSGSVASYSRTHL  
CDVEIPGGQPMRESNYTKPPKTPAGKVLGATCYDMRFPPELSKLQAQAEI  
LTYSAFGSVTPGAWHEVLLRARAIQSCYIAAAQCGRRHETRAYSGHSMVDPWGT  
VVARCSEGPGLCLARIDLHFLQOMROHLPVQHRRPDLVSGLSGLHPLS"  
BASE COUNT 368 a 334 c 331 g 332 t  
ORIGIN  
alignment\_scores:  
Quality: 1476.00 Length: 327  
Ratio: 4.824 Gaps: 2  
Percent Similarity: 93.578 Percent Identity: 84.098  
alignment\_block:  
US-09-357-675C-21 x BC021634 ..  
Align seg 1/1 to: BC021634 from: 1 to: 1365  
1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17  
|||||  
|||||

59	ATGCTGGGCTTCATCAGCAGGCTCCTCACCAA.....CTCTGTGTG	99
17	sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgp	34
100	TACCGGATACCGATTACTTCGNACCCAGTAGTTGTACTACGCCAGGC	149
34	roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla	50
150	CCAGAACCATGTCC...TCATCAACTTCCTGGAGCTGCCCTGGTGGCT	196
51	valCysGlnValThrSerThrProAspLysGlnGlnAsnPhelysThrCy	67
197	GTGTGCCAGGTAAACATCAACACCAACAAACAGACAGAACTTTAAACATG	246
67	sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL	84
247	TGCTGAGTTGGTTCAAGAGGCTGCCAGACTGGGTGCTTGCTGGCCTTTC	296
84	euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis	100
297	TGCCTGAGGCATTGACTTTATTGCACGAAACCCCTGCCGAGACATTACTC	346
101	LeuSerGluProLeuGlyLysLeuLeuGluGluTyrThrGlnLeuAl	117
347	CTGTCCGAACCACTGAATGGGGGATCTTTTGGGCCAAATTACGACAGCTTG	396
117	aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG	134
397	CAGGGNAATGTGAATCTGGCTGTCTCTGGGGCGTTTCCAGACGCTGGCC	446
134	lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn	150
447	AAGACTGGCGAGCAGAATCAGAAAAATCTACAAATTGTCATGTGCTTTGAAC	496
151	SerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspVa	167
497	AGCAAGGGATCAGTAGTGGCCAGTTTACAGGAAGACACATCTGTCCGATGT	546
167	lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG	184
547	AGAGATCCCAAGCTCAGGGGCCGATGAGAGAAGCACTATACCAAGCCTG	596
184	lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu	200
597	GAGGCACCTCTCAGCCACCTGTCTAAGACACCGCGCTGCCAAGTTGGTCTA	646
201	AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl	217
647	GCAATCTGTTATTGACATCGCGTTCCTCGAACTTTCTTTGAAATTGGCTCA	696
217	nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG	234
697	AGCTGGGGCAGAAATPACTTACTTCTTCAGCCCTTTGGATCTGTTACAG	746
234	lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln	250
747	GTCCAGCCCACTGGGAGGTGCTGTCCGGGGCCCGCCCATTTGAATCTCAG	796
251	CysTyrValValAlaAlaAlaGlnCysGlyArgHisGluLysArgAl	267
797	TGCTATGTATATAGCAGCAGCAGCTGTGTGACGCCCAACCATGAAACAAGAGC	846
267	aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAla	284
847	AGTTATGGCCATACATGTGTGGTTGACCGGTGGGGCACAGTGGTGGCCC	896
284	rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr	300
897	GCTGTCCGAGGACCGGCTCTGCCTTGTCTCGAATTGATCTCCACTTT	946
301	LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgProAs	317
947	CTACAAAGATGCGCAACACTGCTTGTGTTTCCAGACCCGACAGCTGA	996

317 pLeuTyrGLyAsnLeuGLyHisProfeUser 327  
|||||:::|||||  
997 CCTCATGGCAGTCTGGGTATCCACTCTCT 1027

seq\_name: gb\_pr:AF069984

seq\_documentation\_block:  
LOCUS AF069984 . 4079 bp DNA linear PRI 23-JUL-1998  
DEFINITION Homo sapiens nitrlase homolog 1 (NIT1) gene, alternatively spliced  
product, complete cds.  
ACCESSION AF069984  
VERSION AF069984.1 GI:3242977  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 4079)  
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Weimuth,P., Rothman,J.H., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Nitrilase and fhit homologs are encoded as fusion proteins in  
Drosophila melanogaster and caenorhabditis elegans  
Proc.Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
98337986  
2 (bases 1 to 4079)  
Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,  
Tillib,S., Draganescu,A., Weimuth,P., Rothman,J., Huebner,K.,  
Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.  
Direct Submission  
Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA  
FEATURES  
Source  
1. 4079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q21-q22"  
join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3569)  
/gene="NIT1"  
/note="alternatively spliced"  
/product="nitrlase homolog 1"  
635..3569  
/gene="NIT1"  
635..680  
/gene="NIT1"  
/note="alternatively spliced"  
/number=1  
join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,  
2609..2734,3001..3267)  
/gene="NIT1"  
/note="NIT1; alternatively spliced"  
/codon\_start=1  
/product="nitrlase homolog 1"  
/protein\_id="AAC39901.1"  
/db\_xref="GI:3242978"  
/translation="MLGIFRPHRFLSLCPGLRIPQLSVLCQAPRPMIAISSSS  
ELPLVAVQVTPDKQONPKTCAELVREARALGACLAPLPAEDFTARDPAETLHLS  
EPLGKLLERYTOLARCGLWLSGGFHGQGDWEOTKLYNCHVLLNKGAVVATYR  
KTHLCDEIPIGQCPMCESNTPGCSLESPTAGKIGLAVCYDMRPELSLALAOA  
CAEILLYPSAGSGITGPAAHEVLLARAEITOCYVVAACQCRHKEKRASIGHSMVVD  
PWGTVVARCGPGLCLARIDLNYLRQLRHLPLVFQHRRLPDLYNGLHPLS"  
681..711  
/gene="NIT1"  
/note="1A; alternatively spliced"  
712..741  
/gene="NIT1"  
/note="1B; alternatively spliced"  
940..1053  
/gene="NIT1"  
/note="1C; alternatively spliced"

mRNA

gene

exon

CDS

exon

exon

exon



2689 CTATCCTTCAGCTTTTGGATCCATTACAGGCCAGCCACCTGGGAGGTAA 2738  
241 Leu..... 241  
2739 GATGATGCCCTTTTAAACATAAGGGCTTTTCTTAACCTCATCTCCCC 2788  
241 ..... 241  
2789 CCTTGGCCCTACCAGTTAAATTCCTTCCCTTCCACCTAATGGGAAAC 2838  
241 ..... 241  
2839 TCATTCGCCAGATATTTCTCTCATGATAGTTAAATAGTAAATCATTC 2888  
241 ..... 241  
2889 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCCC 2938  
241 ..... 241  
2939 TCACCTGTCACTTCCCCACATTTTGTCTACATGTACTTAAAGTGAACACACA 2988  
242 ..... LeuArgAlaAraArgAlaAlaGluThrGlnCysTy 252  
2989 TCTCATGCCAGGTGTTGCTGCGGGCCGCTGCTATCGAAACCCAGTGCTA 3038  
252 rValValAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269  
3039 TGTAGTGGCAGCAGCAGTGTGGAGCCACCATGAGAGAGAGCAAGTT 3088  
269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaAraGys 285  
3089 ATGCCACACAGTGTGTAGACCCCTGGGGAACAGTGGTGGCCGCTGC 3138  
286 SerGluGlyProGlyLeuCysLeuAlaAraGlnLeuAsnTyrLeuAr 302  
3139 TCTGAGGGGCGAGCCCTCTGCTTGGCCGAATAGACCTCAACTATCTGCG 3188  
302 gGlnLeuArgHisLeuProValPheGlnHisArgArgProAspLeuT 319  
3189 ACAGTTCGGCGGACACCTGCTGTGTCCAGACCCGAGCCCTGACCTCT 3238  
319 yrGlyAsnLeuGlyHisProLeuSer 327  
3239 ATGGCAATCTGGGTCAACCCACTGTCT 3264

seq\_name: gb\_htg:AL354714\_4

seq\_documentation\_block:

WPCOMMENT  
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714  
Fragment Name Begin End  
AL354714\_0 1 110000  
AL354714\_1 100001 210000  
AL354714\_2 200001 310000  
AL354714\_3 300001 410000  
AL354714\_4 400001 510000  
AL354714\_5 500001 610000  
AL354714\_6 600001 699820

Continuation (5 of 7) of AL354714 from base 400001 (AL354714 Homo sapiens chromosome 1)

alignment\_scores:  
Quality: 1468.50 Length: 660  
Ratio: 4.505 Gaps: 8  
Percent Similarity: 49.394 Percent Identity: 48.939

alignment\_block:

US-09-357-675C-21 x AL354714\_4 ..

Align seg 1/1 to: AL354714\_4 from: 1 to: 110000

2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18

103858 CTGGGCTTCATCACCAGGCTCTCTCAGATTCTGTGCTCCCTTCTGTGTC 103907  
18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32  
103908 TGGACTCCGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGTAAC 103957  
32 ..... 32  
103958 ACGTTTTGTGTCTCCTCAGTGCCTGCACCTTAGATGCTCAGTTTGTAA 104007  
32 ..... 32  
104008 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATTGGT 104057  
32 ..... 32  
104058 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAAGAGCATGATCAAAAG 104107  
32 ..... 32  
104108 GAAGTCCAGCTTTCTTCCCTCTCCACTTGCACCCCTTAGCATTAAATTTG 104157  
33 ..... ArgPr 34  
104158 CTTCCCTGTGCTATGAAATCTGAGAAATCTGCCTATGCTCTTCCACAGGCC 104207  
34 oArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAlaV 51  
104208 CAGAGCCATGGCTATCTCCTCTTCTCCGAACTGCCCTGGTGGCTG 104257  
51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67  
104258 TGTGCCAGGTAACTCGACGCCACACAGCAAGCACTTTAAACATGT 104307  
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84  
104308 GCTGAGCTGGTTCGAGAGGCTGCAGACTGGGTGCCTGCTGCTTCTCCT 104357  
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101  
104358 GCCTGAGGCATTTGACTTTCATTCACGGGACCTGCAGAGACGCTACACC 104407  
101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla 117  
104408 TGTCTGAACCACTGGGTGGGAAACTTTTGGAGAATACACCCAGCTTGC 104457  
117 ..... 117  
104458 AGGTATCAGGGAATAGCGAGGAGAGGTAGATACTTTTGTGGACAGTGT 104507  
117 ..... 117  
104508 CCCTGGGTTGCCAGATATAGGGTAGAGCCTTCGAGAAGTCAGTGAAGAGT 104557  
118 ..... ArgGluCysGlyLeuTrpLeuSerLeuGly 127  
104558 TGTCACTGTCCCTTCCCCCAGGGAATGTGGACTCTGGCTGCTCTGGGT 104607  
128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144  
104608 GGTTCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAAATCTACAA 104657  
144 nCysHisValLeuLeuAsnSer...Lys..... 152  
104658 TTGTACGTGCTGCTGAACGAAAGGTGAGACTTTATAACCCCTTAGCC 104707  
152 ..... 152  
104708 TGCCTCTTCCCCTGCTTCTTCTACCTAGATTCTCCAGAAATGTTTCTCAAC 104757  
152 ..... 152

```
104758 TCCTATTTCCTTGACCCCAAGGATTTAGGGGTGGTCTCTACTTCAAGTTCCCTA 104807
152 ..... 152
104808 GCCTATAAACTATCTCCTCCTTGGAGGAGTAAGCAAGGCTTCTAGAACAA 104857
153 .....GlyAlaValAlaThrT 159
104858 CCAGCACTGATATTCCTTCTTCTTACTGTAGGGCAGTAGTGCCCACTT 104907
159 yrArgLysThrHisLeuCysAspValGlulileProGlyGlnGlyProMet 175
104908 ACAGGAAGACACATCTGTGTGACCTGTAGAGATTCAGGGCCAGTTCAGTGTG 104957
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
104958 TGTGAAAGCAACTCTACCATGCCTGGGCCAGTCTTGAGTCCACCTGTGAG 105007
192 rThrProAlaGlyLysIleGly_Leu..... 200
105008 CACACCAGCAGGCAAGGTAGGATTGTGAAAGGATGAGGGAGGGGAACAG 105057
200 ..... 200
105058 GAATACTTTGAACTGGCAGTAGAGATAGAAAGCCCTTAAGAGAGGGGTA 105107
200 ..... 200
105108 ATGGAATATGACTAGATGCTGTGACAAACAGACAGGAGGAACTACTAAG 105157
201 .....AlaValCysTyrAspMetArg 207
105158 TAGGCTGTTTTTCATTCAGATGGTGTAGCTGTCTGCTATGACATGCGG 105207
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGlulileuTh 224
105208 TTCCTGAACTCTCTCGGGATTGGCTCAAGCTGGAGCAGAGATACTTAC 105257
224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal_ 240
105258 CTATCCTTCAGCTTTTGATCCATTACAGGCCAGCCCACTGGAGAGTAA 105307
241 Leu..... 241
105308 GATGATGCCTTTTTAAACATAAGGGCCTTTTCTTAACCTCATCTTCCCC 105357
241 ..... 241
105358 CCTTGGCCCTACCAGTTAAATTCCTTCCCTTTTCCACCTAATGGGAAAC 105407
241 ..... 241
105408 TCATTCCCCAGATATTCTCTCATGAATAGTTAAATAGTAATCATCTCC 105457
241 ..... 241
105458 TAGGCAATTACCAAAATAGTCACAAATGGGTAGATTGGTCTGTAATGTCCC 105507
241 ..... 241
105508 TCACCTGTCTACTCCCCCACTATTGTTGTACATGTAAGTGAACACACA 105557
242 .....LeuArgAlaArgAlaIleGluThrGlnCysTy 252
105558 TCTCATGCCCGAGGTGTGCTGGGCCCGTGTCTATCGAAACCCAGTGCTA 105607
252 rValValAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269
105608 TGTAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGCAAGTT 105657
269 yrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285
105658 ATGCCACAGCATGGTGTAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 105707
```

```
286 SerGluGlyProGlyLeuCysLeuAlaAraGileAspLeuAsnTyrLeuAr 302
105708 TCTGAGGGCCAGGCTCTGCTTGGCCGAATAGACCTCAACTATCTGCG 105757
302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319
105758 ACAGTTGGCCGACACCTGCTGTGTTCAGCACCGCAGGCTGACCTCT 105807
319 yrGlyAsnLeuGlyHisProLeuSer 327
105808 ATGGCAATCTGGGTCCACCCACTGTCT 105833
seq_name: gb_htg:AL354714_5
seq_documentation_block:
WPCOMMENT
Sequence split into 7 fragments LOCUS AL354714 Accession AL354714
Fragment Name Begin End
AL354714_0 1 110000
AL354714_1 100001 210000
AL354714_2 200001 310000
AL354714_3 300001 410000
AL354714_4 400001 510000
AL354714_5 500001 610000
AL354714_6 600001 699820
Continuation (6 of 7) of AL354714 from base 500001 (AL354714 Homo sapiens chromosome
alignment_scores:
Quality: 1468.50 Length: 660
Ratio: 4.505 Gaps: 8
Percent Similarity: 49.394 Percent Identity: 48.939
alignment_block:
US-09-357-675C-21 x AL354714_5 ..
Align seg 1/1 to: AL354714_5 from: 1 to: 110000
2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
3858 CTGGGCTTCATCACACAGGCTCCTCACAGATTCTCTCTCTCTGTCTGTC 3907
18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
3908 TGGACTCCGGATACCTCACTCTCAGTACTTTGTGCTCAGCCCAAGTAAC 3957
32 ..... 32
3958 ACGTTTGTGTGTCCTCAGTGGCTGGCACTTAGATGCTCAGTTTGTAA 4007
32 ..... 32
4008 ATGGATAGTGGGAGACACAGGAGTGTCAACTATCCACACATTTGATGGT 4057
32 ..... 32
4058 GAGCCCTACTAGCCCTGGGTCAAGTCCCTGTAAAGAGCATGATCAAAAG 4107
32 ..... 32
4108 GAAGTCCAGCTTTCTGCTCTCCACTTGCACCTTGCACCTTAGCATTTAAATTTG 4157
33 .....ArgPr 34
4158 CTTCCCTGTCTATGAATCTGAGATCTTCGCTATGCTGTTCACAGGCC 4207
34 oArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAlav 51
4208 CAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4257
51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
|||||
```

```
4258 TGTGCCAGGTAAACATCGACCCACAGCAAGCAACAGCAACTTTTAAACATGT 4307
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
|||||
4308 GCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGGCTTCCT 4357
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
|||||
4358 GCCTGAGGCATTTGACTTCATTCGACGGGACCCTGCAGAGACGCTACACC 4407
101 euSerGluProLeuGlyGlyLysLeuLeuGluGlyThrGlnLeuAla 117
|||||
4408 TGTCTGAACCACTGGGTGGGAACTTTTGAAGAATACACCCAGCTTGCC 4457
117 ..... 117
4458 AGGTATCAGGGAATAGCAGGAGGAGAGGTAGAATCTTTGTTGGACAGTGT 4507
117 ..... 117
4508 CCCTGGGTTGCCAGATATGAGGCTAGAGCCTTCAGAAGTCAGTGAAGAGT 4557
118 ..... ArgGluCysGlyLeuTrpLeuSerLeuGly 127
|||||
4558 TGTCAAGTGTCCCTTCCCCCAGGGAATGTGGACTCTGGCTGTCTTGGGT 4607
128 GlyPheHisGluArgGlyGlnAspTTPGluGlnThrGlnLysIleTyraS 144
|||||
4608 GGTTCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAAATCTACAA 4657
144 nCysHisValLeuLeuAsnSer...Lys..... 152
|||||
4658 TTGTCAAGTGTCTCTGAACACGAAGAGTGAGACTTTATAACCCCTTTAGCC 4707
152 ..... 152
4708 TGCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGAAATTGTTTCTCAAC 4757
152 ..... 152
4758 TCTATTTCCTTGACCCAGGATTAGGGGTGCTCTACTTCAAGTTCCTA 4807
152 ..... 152
4808 GCCTATAAACTATCTCTCCTTGGGAGAGTAAGCAAGGCTTCTAGAACA 4857
153 ..... GlyAlaValAlaIleThr 159
|||||
4858 CCAGCACTGATATCTCTTCTTCTACTGTAGGGGCAGTAGTGGCCACTT 4907
159 YrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175
|||||
4908 ACAGGAAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCCTATG 4957
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValse 192
|||||
4958 TGTGAAAGCAACTCTACCATGCTGGGCCCGCAGCTTGAGTCACTGTGAC 5007
192 rThrProAlaGlyLysIleGly...Leu..... 200
|||||
5008 CACACCAGCAGCAAGGTAGGAGTTGTGAAGATGAGGGAGGGGAACAG 5057
200 ..... 200
5058 GAATACTTTGAACCTGGCAGTAGAGGTAGAAACCCCTAAGAGAGGGGTA 5107
200 ..... 200
5108 ATGAAATATGACTAGATGCTGTGACAAACAGACAGGAAGACTACTAAG 5157
201 .....AlaValCysTyraSPMetArg 207
|||||
5158 TAGGCTGTTTTTCATCCAGATTGGTCTAGCTGTCTGCTATGACATGCGG 5207
```

```
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
|||||
5208 TTCCCTGAACCTCTCTCGCATTTGGCTCAAGCTGGAGAGAGATACTTAC 5257
224 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal.. 240
|||||
5258 CTATCCTTCAGCTTTTGGATCCATTACAGGCCAGCCCACTGGGAGTAA 5307
241 ..Leu..... 241
5308 GATGATGCCTTTTAAACATAAGGGCCTTTCTTAACCTCATCTTCCCC 5357
241 ..... 241
5358 CCTTGGCCCTACCAGTTAAATTCCTTCCCCTTTCCACCTAATGGGAAAC 5407
241 ..... 241
5408 TCATTCCCAGATATTTCTCTCATGAATAGTTAAAATAGTAAATCATTC 5457
241 ..... 241
5458 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCC 5507
241 ..... 241
5508 TCACCTGTCACTTCCCACATATTTGCTACATGTACTTAAGTGAACACACA 5557
242 .....LeuArgAlaIleGluThrGlnCysTy 252
|||||
5558 TCTCATGCCAGGTGTTGCTGGGGCCGCTGTATCGAAAACCCAGTGCTA 5607
252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysGluArgAlaSerT 269
|||||
5608 TGTAGTGGCAGCAGCACAGTGTGGAGCCACCATTGAGAAGAGAGCAAGTT 5657
269 YrGlyHisSerMetValValAspProTyrGlyThrValValAlaIleArgCys 285
|||||
5658 ATGCCACAGCATGGTGTAGACCCCTGGGAACAGCTGGTGGCCCGCTGC 5707
286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302
|||||
5708 TCTGAGGGCCAGGCCTCTGCCTTGGCCGAATAGACCTCAACTATCTGG 5757
302 gGlnLeuArgArgHisLeuProValPheGlnHisArgAlaArgProAspLeuT 319
|||||
5758 ACAGTTGCCCGCACACCTGCTGTGTCCAGCACCGCCGCTGACCTCT 5807
319 YrGlyAsnLeuGlyHisProLeuSer 327
|||||
5808 ATGGCAATCTGGGTCAACCCACTGTCT 5833
```

seq\_name: gb\_htg:AL590651

seq\_documentation\_block:

LOCUS AL590651 167863 bp DNA linear HTG 18-AUG-2001  
DEFINITION Homo sapiens chromosome 1 clone RP11-137A12, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 9 unordered pieces.

ACCESSION AL590651

VERSION AL590651.4 GI:13990236

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 167863)

AUTHORS Harrison, E.

TITLE Direct Submission

JOURNAL Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerquest@sanger.ac.uk

COMMENT On May 7, 2001 this sequence version replaced gi:13751001.



```

misc_feature          fragment_chain:2"
137865..163160
/note="assembly_fragment:00117
fragment_chain:2"
163261..167863
/note="assembly_fragment:01267
fragment_chain:2
clone_end:SP6
vector_side:right"
BASE COUNT   43927 a 39935 c 40002 g 43197 t 802 others
ORIGIN

alignment_scores:
  Quality: 1468.50      length: 660
  Ratio: 4.505          Gaps: 8
Percent Similarity: 49.394 Percent Identity: 48.939

alignment_block:
US-09-357-675C-21 x AL590651/rev ..

Align seg 1/1 to reverse of: AL590651 from: 1 to: 167863

2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysPr 18
|||||
151392 CTGGGCTTCATCACCAGGCTCTCACAGATTCCTGTCCCTCTGTGTGCC 151343
|||||
18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
|||||
151342 TGGACTCCGGATACCTCAACTCTCAGTACTTGTGCTCAGCCCAAGTAAC 151293
|||||
32 ..... 32

151292 ACGTTTGTGTGTCCTCAGTGCTGGCACTTAGATGCTCAGTTGTATA 151243
|||||
32 ..... 32

151242 ATGGATAGTGGGAGACACAGGAGTGCACTATCCACACATTCATGATGGT 151193
|||||
32 ..... 32

151192 GAGCCCTACTAGCCCTGGGTCAAGTGCCTGTAAAGAGCATGATCAAAAG 151143
|||||
32 ..... 32

151142 GAAGTCCAGCTTTCCTGCCTCTCCACTTGCACCCCTTAGCATTAATTTG 151093
|||||
33 .....ArgPr 34
|||||

151092 CTTCCTGTGCTATGAATCTGAGAATCCTGCCTATGCTGTTTCACAGCC 151043
|||||
34 oArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAlaV 51
|||||
151042 CAGAGCCATGGGTATCTCCTCTTCCTCCTCGGAACCTGCCCTGTGGTG 150993
|||||
51 aCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67
|||||
150992 TGTGCCAGGTACATCGAGCCACAGCAAGCAAGCAACTTTAAACATGT 150943
|||||
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
|||||
150942 GCTGAGCTGTTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTCT 150893
|||||
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101
|||||
150892 GCCTGAGGCATTTGACTTCAATTCACGGGACCCCTGCAGACGCGTACAC 150843
|||||
101 euSerGluProLeuGlyGlyLysLeuLeuGluGluTyThrGlnLeuAla 117
|||||
150842 TGTCTGAACCACTGGGTGGGAACTTTTGGGAAGAAATACACCCAGCTGCC 150793
|||||
117 ..... 117

```

150792 AGTATCAGGAAATAGCGAGGAGAGGTAGATCTTTGTTGACAGTGT 150743  
117 ..... 117  
150742 CCCTGGTTGCCAGATATGAGGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 150693  
118 ..... ArgGluCysGlyLeuTrpLeuSerLeuGly 127  
150692 TGTCAAGTGTCCCTTCCCCCAGGAGATGTGGACTCTGGCTGTCTGGGT 150643  
128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAs 144  
150642 GGTTCATCAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAAAATCTACAA 150593  
144 nCysHisValLeuLeuAsnSer...Lys..... 152  
150592 TTGTCAAGTGTCTGTAACAGCAAGAGGTGAGACTTTATAACCCCTTAGCC 150543  
152 ..... 152  
150542 TGCCCTCTCCATGCTCTTCTACCTAGATTCTCCAGAAATTGTTCTCAAC 150493  
152 ..... 152  
150492 TCTTATTCTCTTGACCCAAAGATTAGGGGTGTCTACTTCACTTCTCCTA 150443  
152 ..... 152  
150442 GCCTATAAACTATCTCTCTCTGGGAGGAGTAGCAAGGCTTCTAGAAC 150393  
153 ..... GlyAlaValAlaIleThrT 159  
150392 CCAGCACTGATATCTCTTCTTACTGTAGGGGAGTAGTGGCCACTT 150343  
159 YrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175  
150342 ACAGGAAGACACATCTGTGTGACGTAGATTCAGAGATTCCAGGGCAGGGCCATG 150293  
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192  
150292 TGTGAAGAACAATCTACCATGCTCTGGCCGAGTCTTGAGTCACTGTGAC 150243  
192 rThrProAlaGlyLysIleGly...Leu..... 200  
150242 CACACCAGCAGGCAAGGTAGGAGTGTGAAAGTCAGGGGAGGAGACAG 150193  
200 ..... 200  
150192 GAATACTTTGAAGTGGCAGTAGAGGATAGAAAGCCCTAAGAGAGGGGTA 150143  
200 ..... 200  
150142 ATGGAAATATGACTAGATGCTGTGACAAACAGACAGAGGAGACTACTAAG 150093  
201 ..... AlaValCysTyrAspMetArg 207  
150092 TAGGCTGTTTTTCATTCAGATTGGTCTAGCTGTCTGCTATGACATCGCG 150043  
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224  
150042 TTTCCCTGAATCTCTCTGGCATTGGCTCAAGCTGGAGCAGAGATACTTAC 149993  
224 rTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240  
149992 CTATCCTTCAGCTTTTGGATGCCATTACAGGCCAGCCCACTGGGAGGTAA 149943  
241 ..... 241  
149942 GATGATGCCTTTTAAACATAGGGCCCTTTCTTAACCTCATCTTCCCC 149893  
241 ..... 241  
149892 CCTTGGCCCTACCAAGTTAAATCTTCCCTTCCACCTTAATGGGAAAC 149843

241 ..... 241  
149842 TCATTCCCCAGATATTTCTCTCATGAATAGTTAAATAGTAATCATTC 149793  
241 ..... 241  
149792 TAGGCAATTACCAAAATAGTCACAATGGGTAGATTGGTCTGTAATGTCC 149743  
241 ..... 241  
149742 TCACCTGTCACTTCCCCACTATTGCTACATGTACTTAAGTGAACACACA 149693  
242 ..... LeuArgAlaArgAlaIleGluThrClnCysty 252  
149692 TCTCATGCCAGGTGTGTGGGGCCCTGCTATCGAAACCCAGCGTCTA 149643  
252 rValValAlaAlaGlnCysGlyArgHisGlnLysArgAlaSerT 269  
149642 TGTAGTGGCAGCAGCACAGTGTGGAGCCACCATTGAGAAGAGAGCAAGT 149593  
269 YrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285  
149592 ATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGGTGGCCGCTGC 149543  
286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302  
149542 TCTGAGGGCCAGGCTCTGCTTGGCCGAATAGACCTCACTATCTCG 149493  
302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuT 319  
149492 ACAGTTGCCGCCGACACCTGCCTGTGTTCAGCACCGCAGCCTGACCTCT 149443  
319 YrGlyAsnLeuGlyHisProLeuSer 327  
149442 ATGGCAATCTGGGTCAACCACCTGTCT 149417

seq\_name: gb\_pr:AL591806

seq\_documentation\_block:

LOCUS AL591806 208822 bp DNA linear PRI 30-JAN-2002  
DEFINITION Human DNA sequence from clone RP11-544M22 on chromosome 1, complete  
sequence.

ACCESSION AL591806

VERSION AL591806.16 GI:18476709

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 208822)

AUTHORS Harrison,E..

TITLE Direct Submission

JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerv@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 1, 2002 this sequence version replaced gi:17902927.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em:, EMBL; SW:,

SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP

database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>  
 RP11-544M22 is from the library RPCI-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-544M22. The true  
 left end of clone RP11-137A12 is at 156538 in this sequence. The  
 true right end of clone RP11-381D2 is at 145015 in this sequence.

#### FEATURES

Source  
 1. .200822  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-544M22"  
 /clone\_lib="RPCI-11.2"  
 10395..10397  
 misc\_feature  
 /note="Sequence from overlapping clone RP11-381D2  
 (AL162592). Assembly confirmed by restriction digest."  
 44098..44107  
 misc\_feature  
 /note="Sequence from overlapping clone RP11-381D2  
 (AL162592). Assembly confirmed by restriction digest."  
 59585..59604  
 misc\_feature  
 /note="Sequence from overlapping clone RP11-381D2  
 (AL162592). Assembly confirmed by restriction digest."  
 90797..90883  
 misc\_feature  
 /note="Single clone region. Reads generated from a  
 transposon library derived from a single pUC clone.  
 Restriction digest data confirm the assembly."  
 90861..90883  
 misc\_feature  
 /note="Sequence from uni-directional dGTP big dye  
 terminator reads only."  
 55215 a 49052 c 46183 g 50372 t

#### alignment\_scores:

Quality: 1468.50 Length: 660  
 Ratio: 4.505 Gaps: 8  
 Percent Similarity: 49.394 Percent Identity: 48.939

#### alignment\_block:

US-09-357-675C-21 x AL591806 ..

Align seg 1/1 to: AL591806 from: 1 to: 200822

2 LeuGlyPheLeThrArgProHisArgPheLeuSerLeuLeuCysPr 18  
 |||||  
 172789 CTGGGCTTCATCACAGCCCTCCACAGATTCTGTCCTTCGTGTC 172838  
 18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32  
 |||||  
 172839 TGGACTCGGATACCTCAACTCTCAGTACTTGTGCTCAGCCCAAGTAAC 172888  
 32 ..... 32  
 172889 ACGTTTGTGTCCTCAGTGCCTGGCAGCTTAGATGCTCAGTTGTAA 172938  
 32 ..... 32  
 172939 ATGGTAGTGGGAGACACAGAGTGTCAACTATCCACACATTTGATGGT 172988  
 32 ..... 32  
 172989 GAGCCCTACTAGCCCTGGGTCAAGTGCCTGTAAAGACATGATCAAAAG 173038  
 32 ..... 32  
 173039 GAAGTCCAGCTTCTGCTCTCCACTTGCACCCCTTAGCATTAAATTG 173088

33 .....ArgPr 34  
 |||||  
 173089 CTTCCCTGCTATGAATCTGAGATCCCTGCCTATCTGTTACAGGCC 173138  
 34 oArgAlaMetalalleSerSerSerCysGluLeuProLeuValAlav 51  
 |||||  
 173139 CAGAGCCATGGCTATCTCTCTCTCGAATCCCTGCTGCTGCTGCTG 173188  
 51 aCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCys 67  
 |||||  
 173189 TGTGCCAGGTAAATCAGCCAGCAAGCAAGCAACATTTAAACATGT 173238  
 68 AlaGluLeuValArgGluAlaAlaAArgLeuGlyAlaCysLeuAlaPheLe 84  
 |||||  
 173239 GCTGAGCTGTTTCAGAGAGCTGCCAGACTGGGTGCCTGCTGCTGCT 173288  
 84 uProGluAlaPheAspPheLeAlaArgAspProAlaGluThrLeuHisL 101  
 |||||  
 173289 GCCTGAGGCATTTGACTTCATTGCACGGGACCTGCAGAGACGCTACAC 173338  
 101 euSerGluProLeuGlyClyLysLeuLeuGluGluTyThrGlnLeuAla 117  
 |||||  
 173339 TGTCTGAACCACTGGGTGGGAACTTTTGGAGAATACACCAAGCTTGCC 173388  
 117 ..... 117  
 173389 AGGTATCAGGGAATACCGAGGGAGAGGTAGAATCTTTGTTGGACAGTGT 173438  
 117 ..... 117  
 173439 CCCTGGCTTCCAGATATGAGGTAGAGCCTTGAGAAGTCAGTGAAGAGT 173488  
 118 .....ArgGluCysGlyLeuTrpLeuSerLeuGly 127  
 |||||  
 173489 TGTCAAGTGTCCCTTCCCCCAAGGAATGTGGACTCTGGCTGCTGGGT 173538  
 128 GlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyAs 144  
 |||||  
 173539 GGTTCATGAGCTGGCCAGAGCTGGGAGCAGACTCAGAAATCTACAA 173588  
 144 nCysHisValLeuLeuAsnSer..Lys..... 152  
 |||||  
 173589 TTGTCAAGTGTGCTGAACAGCAAAAGGTGAGACTTTATAACCCCTTAGCC 173638  
 152 ..... 152  
 173639 TGCCTCTCCCATGCTCTTCTACCTAGATTCTCCAGAAATTGTTTCTCAAC 173688  
 152 ..... 152  
 173689 TCTTATTTCTTGACCCCAAGGATTTAGGGGTGGTCTCTACTTCAGTTCTTA 173738  
 152 ..... 152  
 173739 GCCTATAAAGTATCTCTCTCTGGGAGGAGTAAAGAGCTTCTAGAACA 173788  
 153 .....GlyAlaValAlaAlaThrT 159  
 |||||  
 173789 CCAGCACTGATATTCTCTTCTTACTGTAGGGGCAGTAGTGGCCACTT 173838  
 159 yArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175  
 |||||  
 173839 ACAGGAAGACACATCTGTGTGACGTAGATATCCAGGGGAGGGCCCTATG 173888  
 176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValse 192  
 |||||  
 173889 TGTGAAGCAACTCTACCATGCCTGGGCCAGTCTTGAGTCAACCTGTAG 173938  
 192 rThrProAlaGlyLysIleGly..Leu..... 200  
 |||||  
 173939 CACACAGCAGGCAAGGTAGGAGTTGTGAAGGATGATGGGAGGGGAACAG 173988

```

SOURCE          house mouse.
ORGANISM        Mus musculus
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          1 (bases 1 to 211771)
JOURNAL        Deschamps, S., Gu, W. and Roe, B. A.
REFERENCE       Mus musculus BAC Clone rp23-395h6
AUTHORS        2 (bases 1 to 211771)
TITLE          Direct Submission
JOURNAL        Deschamps, S., Gu, W. and Roe, B. A.
COMMENT        Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
                The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                OK 73019, USA
                ----- Genome Center
                Center: Department Of Chemistry And Biochemistry
                The University Of Oklahoma
                Center code: UOKNOR
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 1 contigs. Gaps between the contigs
                * are represented as runs of N. The order of the pieces
                * is believed to be correct as given, however the sizes
                * of the gaps between them are based on estimates that have
                * provided by the submitter.
                * This sequence will be replaced
                * by the finished sequence as soon as it is available and
                * the accession number will be preserved.
                * 1 211771: contig of 211771 bp in length.
                Location/Qualifiers
                1. 211771
                   /organism="Mus musculus"
                   /db_xref="taxon:10090"
                   /chromosome="1"
                   /clone="rp23-395h6"
                   /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT     55622 a 48688 c 49207 g 58231 t 23 others
ORIGIN

alignment_scores:
    Quality: 1105.50          Length: 676
    Ratio: 3.893              Gaps: 8
Percent Similarity: 42.012    Percent Identity: 37.722

alignment_block:
US-09-357-675C-21 x AC084821 ..

Align seg 1/1 to: AC084821 from: 1 to: 211771

29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysG1 45
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
975 TGTGTCTTTCACGAGCCAGACCATGTCC...TCATCAACTTCTCTGGGA 1021

45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysG1og 62
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
1022 GCTGCCCTTGGTGGCTGTGTCAGGTACATCAACACCAAGCAAGCAAG 1071

62 InAsnPhelyThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 78
::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1072 AGAACTTTAAACATCTGCTGAGTTGGTTCAAGAGCGTGCCAGACTGGGT 1121

79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPr 95
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
1122 GCTGCCCTGGCCTTTCTGCCTGAGGCATTTGACTTTATTGCAGAAACCC 1171

95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluG 112
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
1172 TGCCGAGACATTACTCTCTGCGAACCACTGAATGGGATCTTTTGGGCC 1221

112 LutyThrGlnLeuAla..... 117
::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||

```

```
1222 AATATGCCAGCTTGCCAGGTACAAGGTTAGGGTTTGGGGGGGGGGGT 1271
117 ..... 117
1272 TGTGTGTAGAACTCCCTGTTGAGCAGTGTCTCTGGATTGCCACAGG 1321
118 ..... ArgGluCysGlyL 122
1322 GGTAGAGTTGTGTGTGTCGTGTCCTCCCTCCCTCCCTCCCTCCCT 1371
122 euTrpLeuSerLeuGlyGlyPheHisIleValLeuAsnSerLeuGlyAla 138
1372 TCTGGCTGCTCTGGGGGTTTCCACAGGCTGGCCCAAGACTGGGAGCAG 1421
139 ThrGlnLysIleValAsnSerLeuValLeuAsnSerLeuGlyAla 154
1422 AATCAGAAAATCTACAAATGTCATGTCCTTTGNACAGCAAGGTTGAGAC 1471
155 Val 155
1472 GTCATAGCCCTTAGTTTGCCTCTTCCAGGCTTCTCTACCTTTCCCTCAAC 1521
155 ..... 155
1522 TCAGATCTCAGCATTTCTTTCTCAACTGCTTGGCCTTGGGGTTTTATGC 1571
155 ..... 155
1572 TTCTGTTCTCAGCCTACAGTTACTCTTTGGGAGAGAGTAAGCTGGAC 1621
155 ..... 155
1622 TTGTAGAGTGTCTACTACTGAGTATTTCTCTCTCTCTCTCTCTCTCTC 1671
156 ..... ValAlaThrTyrA 160
1672 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1721
160 rGlyThrHisLeuCysAspValIleProGlyGlnGlyProMetCys 176
1722 GGAAGACACATCTGCGGATGTAGAGATCCAGGTACGGGCCGATGAGA 1771
177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
1772 GAAAGCAACTATACCAAGCTGGAGGCACTCTTGAGCCACTCTGCAAGAC 1821
193 rProAlaGly 197
1822 ACCGGCTGGCAAGGTGGGATTTTGGAAAGAAATGGGACTATGTTGAAA 1871
197 ..... 197
1872 GTGGTAACACCCCTTGTGTGAGGCTAATGAGGCACCTCAGTGTCTG 1921
198 ..... IleGlyL 200
1922 TTGAATGGACAGGAAAGATGAGTGGGTTGTTTCTCATTTTCAGTTGGTC 1971
200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
1972 TAGCAATCTCTTATGACATCGGTTCCCTGAACTTTCTTTCAATTTGGCT 2021
217 GlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleTh 233
2022 CAAGCTGGGCGAANAATACCTTACTTCTTCCAGCCTTTGGACTGTGTAC 2071
233 rGlyProAlaHisTrp 238
2072 AGTCCAGCCCACTGGGAGGTAAAGAAATGCCCTTCTTCAATATATAGC 2121
238 ..... 238
2122 CTCCTTTAACTTACTTTTAGTAGTGGGAGCAGAGTGGTACAAACCTAAA 2171
```

```
238 ..... 238
2172 TCCAAGCAGCTTCAGTGGATAGCACACAGGTTCAAAGCCAGCATGCCTTGAC 2221
238 ..... 238
2222 TAAAAAAGAAAAACAATAAGAGCTGGCAGTGGGGGTGCACACCTTTAA 2271
238 ..... 238
2272 TCCAGCACTCAGGAGGCACAAGAGGTGATCTCTGAGTTTGAGGCCAG 2321
238 ..... 238
2322 CCTGGTCTACAGAGTGGGTTCCAGGACAGCCAGGGCTACACAGAGAAACC 2371
238 ..... 238
2372 CTGTCTCAAAAAACAAAAAGCAACCCACACCTTCAAAGAGTGGGGCATGG 2421
238 ..... 238
2422 TGGTGATGCTCTTCACTCAGCACCTGTGAGCAAAAGGCAGGAGATCTC 2471
238 ..... 238
2472 TGTGGGTCTAAGCCAGCCAGGGTTACACAGGGAGACCTTATCTCAAAAGT 2521
238 ..... 238
2522 ACAAACACAAACACCTCCTCTTCTTACTTTGATTCTCTCTCTTTTCATCT 2571
238 ..... 238
2572 GGTAGGAAATTCATTTCTTAGTAGTGTGAGTCATGTTAAAGAGTTAG 2621
238 ..... 238
2622 TTATTAATCTAGTTGGTTAGTGAAGAGATTGGCCCCACAGTGCCCGCTT 2671
238 ..... 238
2672 CTTGTCACTTGCCTATTAAACGCTGTATGTTCTGGAGTCAACGACGCTCT 2721
239 ..... Glu 253
2722 CATCCTCAGGTGCTGTGCGGGCCCGCCATTTGAATCTCAGTCTATGT 2771
253 lValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrG 270
2772 ANTAGCAGCAGCGAGTGTGACGCCACCATGAACAAGAGCAAGTTATG 2821
270 lHisSerMetValValAspProThrPglyThrValValAlaArgCysSer 286
2822 GCCATAGCATGGTGGTTGACCCGTGGGGCACAGTGTGGCCGCTGCTCC 2871
287 GluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgG 303
2872 GAGGACAGGCGCTCTGCTTGTCTGCGAATTTGATCTCCACTTTCTACAACA 2921
303 nLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrG 320
2922 GATGCGCAACACCTGCTGTGTTCAGCACCCGACGACCTGACCTCTATG 2971
320 lAsnLeuGlyHisProLeuSer 327
2972 GCAGTCTGGGTCTATCCACTCTCT 2994
```

seq\_name: gb\_ro:AF069985

seq\_documentation\_block:

LOCUS

4481 bp

DNA

linear

ROD 23-JUL-1998

DEFINITION Mus musculus nitrlase homolog 1 (Nt1l) gene, alternatively spliced product, complete cds.

ACCESSION AF069985

VERSION AF069985.1 GI:3242979

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 4481)

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE Pekar'sy, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tilib, S., Draganescu, A., Wermuth, P., Rothman, J.H., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

JOURNAL Nitrlase and fhit homologs are encoded as fusion proteins in drosophila melanogaster and caenorhabditis elegans

MEDLINE Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

REFERENCE 2 (bases 1 to 4481)

AUTHORS Pekar'sy, Y., Campiglio, M., Siprashvili, Z., Druck, T., Sedkov, Y., Tilib, S., Draganescu, A., Wermuth, P., Rothman, J., Huebner, K., Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.

TITLE Direct Submission

JOURNAL Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

FEATURES

Source Location/Qualifiers

1..4481

map=1q21-q23

Join(541..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3944)

/gene="Nt1l"

/note="alternatively spliced"

/product="nitrlase homolog 1"

541..3944

/gene="Nt1l"

541..606

/gene="Nt1l"

/note="alternatively spliced"

/number=1

Join(605..606,1177..1263,1638..1889,2015..2118,2362..2495,2626..2751,3392..3658)

/gene="Nt1l"

/note="alternatively spliced"

/codon\_start=1

/product="nitrlase homolog 1"

/protein\_id="AAC40184.1"

/db\_xref="GI:3242980"

/translation="MLGFTITRPPHQLCTGYRLLRIPVLCTQPRPRMTSSSTSWELPLVAVCOVTPNKQENFKTCAELVQEAARGLAFPEAFDFIARNPAETLLSEPLNGDLLQVY SOLARECGIWLISLGFHERGODWEQOKIYNCHVLLSKSGSVASVYRKLHCDVELPGQPNRESNYTRPGTLEPPVKTGPKVGLAICYDMRPFELSLKLAQAGAEI LTPYSAFSGVTPAHWEVLLRARAIESOCYVIAAAQGRRHETRASVGHSMVYDFWGT VVARSGPGGLCLARIDLHFLQOMRQLPVPQHRPRDLYGSLGHLPLS"

863..986

/gene="Nt1l"

/note="alternatively spliced"

/number=2

1177..1263

/gene="Nt1l"

/number=3

1638..1889

/gene="Nt1l"

/number=4

2015..2118

/gene="Nt1l"

/number=5

2362..2495

/gene="Nt1l"

/number=6

2626..2751

/gene="Nt1l"

exon

BASE COUNT 1101 a 1132 c 1133 g 1115 t

ORIGIN

alignment\_scores:

Quality: 1102.50 Length: 680

Ratio: 3.868 Gaps: 8

Percent Similarity: 41.912 Percent Identity: 37.353

alignment\_block:

US-09-357-675C-21 x AF069985

Align seg 1/1 to: AF069985 from: 1 to: 4481

29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSerCysG1 45

1624 TGTGTCATTACAGGCCAGACCATGTCC...TCATCAACTTCTCTGGGA 1670

45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnG 62

1671 GCTGCCCTGGTGGCTGTGTGCCAGGTAACTCAACACCAACCAAGCAAG 1720

62 lnAsnPhelYsThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 78

1721 AGAAGCTTTAAACATGTGTGAGTTGGTTCAGAGGCTGCCAGACTGGGT 1770

79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPr 95

1771 GCCTGCTGGCTTCTGCTGAGGCATTTGACTTTATTCACGAAACCC 1820

95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluG 112

1821 TGCCGAGACATTACTCTGTGCCAGCACTGAATGGGATCTTTTGGGCC 1870

112 luTyThrGlnLeuAlaArg. Glu..... 119

1871 AATATAGCCAGCTTGCAGGTACAGGGTAGGTTTGGGGGGGGGGGGT 1920

119 ..... 119

1921 TGTGCTGTAGAACTCCCTGTTGAGCAGTGTTCCTGAATTGCCACAGAG 1970

120 .....Cy 120

1971 GGCTAGAGTTGTGTGTGCTGTGTCGTGTCCTCCCCCCCCCCCCAGGAATG 2020

120 sGlyLeuTTPLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTTPG 137

2021 TGAATCTGGCTGCTCTTGGCGGGTTTCCACGAGCGTGGCCAAAGACTGGG 2070

137 luGlnThrGlnLysIleTyAsnCysHisValLeuLeuAsnSerLys... 152

2071 AGCAGAATCAGAAAACTACAATTGTGCTGCTTTTGAACACCAAGGCT 2120

152 ..... 152

2121 GAGAGCTCATAGCCCTTTAGTTTGCCTCTTCCAGGCTTCTTACCTTTC 2170

152 ..... 152

2171 CTCAACTCAGATCTTCAGCATTTCTTCTCAACTGCTTGGCCTTGGGGTT 2220

152 ..... 152

2221 TTATGCTTTTGTTCAGCCTACAAGTTACTCTTTGGGAGAGAGCTAAG 2270

152 ..... 152

2271 CTGGACTTGTAGAGTGTCTACTGAGTATTTTCTTTTCTCTCTCTC 2320



```

*      1      11373: contig of 11373 bp in length
*      *      11474: gap of unknown length
*      *      11474: contig of 32004 bp in length
*      *      43478: gap of unknown length
*      *      43478: contig of 82949 bp in length
*      *      126527: gap of unknown length
*      *      126527: contig of 86577 bp in length.
FEATURES
    Location/Qualifiers
        1..213203
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="1"
            /clone="rp23-191a19"
            /clone_lib="RPCI Mouse BAC Library 23"
BASE COUNT  57858 a 50465 c 49150 g 55425 t 305 others
ORIGIN
alignment_scores:
    Quality: 1083.00      Length: 676
    Ratio: 3.813          Gaps: 9
    Percent Similarity: 42.012      Percent Identity: 37.574
alignment_block:
US-09-357-675C-21 x AC087229 ..
Align seg 1/1 to: AC087229 from: 1 to: 213203
29 CysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysGI 45
||||: ||||||:||||: ||||||:||||: ||||||: ||
48642 TGTGTCATTACAGAGCCCAAGAACATGTC...TCATCACTTCCCTGGGA 48688
45 uLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnG 62
||||: ||||||:||||: ||||||:||||: ||||||: ||
48689 GCTGCCCTGGTGGCTGTGCCAGGTAACTCAACACCAACAAACAG 48738
62 lnAsnPhelYsThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 78
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
48739 AGAACTTTAAACATGTGCTGAGTTGGTTCAGAGGCTGCCAGACTGGGT 48788
79 AlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPr 95
||||: ||||||:||||: ||||||:||||: ||||||: ||
48789 GCTTGCGTGGCTTCTGCGTGAAGGATTTGACTTTATTGACGAAACCC 48838
95 oAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluG 112
||||: ||||||:||||: ||||||:||||: ||||||: ||
48839 TGCGGAGACATTACTCTGTGCCGAACCACTGAATGGGATCTTTTGGGC 48888
112 luTyrThrGlnLeuAlaArg..Glu..... 119
:||||:||||:||||: |||
48889 AATATAGCCAGCTTGCCAGGTACAAGGGTAGGGTTTGGGGGGGGGGTT 48938
119 ..... 119
48939 GTGTGTGTAGAACTCCCTGTTGAGCAGTGTCTCTGATTTGCCACAGAGG 48988
120 .....CysGlyLe 122
||||: ||||||:
48989 GGTAGAGTTGTGTGTGTCGTGTCCTCCCTCCCTCCCTCCCTCCCTCC 49038
122 uTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnT 139
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
49039 CTGGCTGCTCCTGGGCGGTTTCCACGAGCGTGCCAAAGACTGGGAGCAGA 49088
139 hrGlnLysIleTrAsnCysHisValIleuLeuAsnSerLysGlyAla..V 155
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
49089 ATCAGAAAATCTACAATGTGCTGTGTTTGAACACAGAGGGGTGAGACG 49138
155 al..... 155
||
49139 TCATAGCCCTTAGTTGCCCTCTTCCAGGCTTCTCTACCTTTCCCTCAACT 49188
155 ..... 155

```

```

49189 CAGATCTTCAGCATTTCTTTCTCAACTGCTTGGCCTTGGGGTTTATGCT 49238
155 ..... 155
49239 TCTGTTCTCAGCCTACAAGTTACTCTTTGGGAGAAGACTAAGCTGGACT 49288
155 ..... 155
49289 TGTAGAGTGTCTACTAGTATTTTCTCTCTCTCTCTCTCTCTCTCTCT 49338
156 .....ValAlaThrTyrAr 160
||||: ||| ||||| .....
49339 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 49388
160 gLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysG 177
||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49389 GAAGACACATCTGTGCGATGTAGAGATCCAGGTCCAGGGCCGATGAGAG 49438
177 luSerAsnSerThrMetProGlyProSerLeuGluSerProValSerThr 193
||||: ||| ||||| .....
49439 AAAGCAACTATACCAAGCCTGGAGGCACCTCTTGAGCCACCTGTCAAGACA 49488
194 ProAlaGly..Lys..... 197
||||: ||| :||
49489 CCGGCTGGCAAGGTGGGAGTTTGGAAAGGAAATGGGACTACTGTTGAAAG 49538
197 ..... 197
49539 TGGTAACACCCCCCATTTGTCGAGGCTAATGAAGGCACCTCAGTGTCTGT 49588
198 .....IleGlyLe 200
||||: |||||
49589 TGAATGGGACAGGAAGATCAGTGGGTTCTTCTCATTTTCAGGTGGTCT 49638
200 uAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49639 AGCAATCTCTTATGACATGCGGTTCCTGAACCTTCTTTGAAATGGCTC 49688
217 lnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233
||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49689 AGCTGGGCGAGAATACATTACTTATCTCTTACGCTTTGGATCTGTGTACA 49738
234 GlyProAlaHisTrp..... 238
||||: |||||
49739 GGTCCAGCCCACTGGGAGGTAAGAGAATGCCCTTTCATACATATAGGCC 49788
238 ..... 238
49789 TCCTTTAACCTTACTTTTAGTAGTGGGAGCAGAGTGGTACAAACCTTAAAT 49838
238 ..... 238
49839 CCAAGCACTTCAGTGGATAGCCACAGGTTCAAAGCCAGCATGCTTGTACT 49888
238 ..... 238
49889 AAAAAAGAAAACAAATAGAAGCTGGGAGTGGGGGTGCACACCTTTAAT 49938
238 ..... 238
49939 CCCAGCACTCAGGAGGAGAAGCAGGTGGATCTCTGAGTTTGAGGCCAGC 49988
238 ..... 238
49989 CTGGTCTACAGAGTGGGTTCCAGGACAGCCAGGCTACACAGAGAAACCC 50038
238 ..... 238
50039 TGCTCAAAAACAAAAAGCAACCCACACCTTCAAAGAGTGGGGCATGGT 50088
238 ..... 238

```



50089	GGTGCATGCTCTTCTTCATCTCAGCACTGTGAGGCAAAAGGACGACGATCTCT	50138
238	.....	238
50139	GTGGGTCTAAGCCACGACGAGGTTCACAGGAGACCTTATCTCAAAAGTA	50188
238	.....	238
50189	CAAAACACAAACACTCTCTCTTTCTACTTGATTTCCTCTTTTCATCTG	50238
238	.....	238
50239	GTGAGGAAATTCATTCTTCTAGCTAGTGTGAGTCATGTGTTAAAGAGTAGT	50288
238	.....	238
50289	TATTAACCTCTAGTGTGAGAGATGTTGGCCACAGTGGCCCGCTTC	50338
238	.....	238
50339	TTGTCACTTGCTTATTAACGCTCTATGTTCTGGAGTCAACGACGCTCTC	50388
239	....Glu.ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal	253
50389	ATCTCTCAGGTCTCTGCGGCCCGCCCAATTGAATCTCAGTGTATGTA	50438
254	ValAlaAlaLglnCysGlyArgHisGlnLysArgAlaSerTyrGln	270
50439	ATACGACGCGCAGTGTGGACGCCACATGAAACAAGACGAATTATGG	50488
270	YHisSerMetValValAspProTrrpGlyThrValValAlaArgCysSerG	287
50489	CCATAGCATGTGTGTACCGCTGGGCACACAGTGTGGCCCGCTGCTCCG	50538
287	LuGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGln	303
50539	AGGACACGAGGCTCTGCTTGCATGATTCATCTCCATCTTCTACAACAG	50588
304	LeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGln	320
50589	ATGGCCCAACACCTGCTGTGTTCCAGCACCCGACCGACCTGACCTCTATGG	50638
320	YAsnLeuGlyHisProLeuSer 327	
50639	CAGTCTGGGTCACTCACTCTCT 50660	
seq_name: gb_htg:AC105589		
seq_documentation_block:		
LOCUS	AC105589	178418 bp DNA linear HTG 09-JAN-2002
DEFINITION	Rattus norvegicus clone CH230-242B2, *** SEQUENCING IN PROGRESS	
ACCESSION	AC105589	*** 65 unordered pieces.
VERSION	AC105589.1	GI:18092812
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
	Rattus.	
REFERENCE	1 (bases 1 to 178418)	
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,	
	Alsbrooks,S.L., Anaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,	
	Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,	
	Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,	
	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,	
	Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,	
	Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,	
	Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,	
	Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Deigado,O.,	
	Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,	
	Dugan-Rocha,S., Durbin,K.J., Earnhart,K.J., Edgar,D., Edwards,C.C.,	
	Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,	

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.B., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojubenok,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,F., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 178418)  
Worley,K.C.  
Direct Submission  
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GNGI  
Center clone name: CH230-242B2  
----- Summary Statistics  
-----  
findPhrapList  
Assembly program: Phrap; version 0.990329First call to  
Consensus quality: 132588 bases at least Q40  
Consensus quality: 14233 bases at least Q30  
Consensus quality: 152072 bases at least Q20  
Estimated insert size: 141845; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-1p estimation  
Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 65 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 8302: contig of 8302 bp in length  
\* 8402: gap of unknown length  
\* 8403  
\* 18472: contig of 10070 bp in length  
\* 18473  
\* 18572: gap of unknown length  
\* 18573  
\* 24922: contig of 6350 bp in length  
\* 24923  
\* 25022: gap of unknown length  
\* 25023  
\* 30899: contig of 5877 bp in length  
\* 30900  
\* 30999: gap of unknown length  
\* 31000  
\* 38320: contig of 7321 bp in length  
\* 38321  
\* 38420: gap of unknown length

```
* 38421 43244: contig of 4824 bp in length
* 43344 43344: gap of unknown length
* 43345 43228: contig of 4884 bp in length
* 48329 48328: gap of unknown length
* 48329 48329: contig of 6257 bp in length
* 54596 54595: gap of unknown length
* 54596 58551: contig of 3856 bp in length
* 58552 58551: gap of unknown length
* 58552 62204: contig of 3553 bp in length
* 62205 62204: gap of unknown length
* 62305 62305: contig of 4283 bp in length
* 66588 66587: gap of unknown length
* 66588 70928: contig of 4241 bp in length
* 70929 71028: gap of unknown length
* 71029 75627: contig of 4599 bp in length
* 75628 75727: gap of unknown length
* 75728 78727: contig of 2545 bp in length
* 78728 78372: gap of unknown length
* 78373 80301: contig of 1929 bp in length
* 80302 80401: gap of unknown length
* 80402 83675: contig of 3274 bp in length
* 83676 83775: gap of unknown length
* 83776 87173: contig of 3398 bp in length
* 87174 87273: gap of unknown length
* 87274 90196: contig of 2923 bp in length
* 90197 90296: gap of unknown length
* 90297 92065: contig of 1769 bp in length
* 92066 92165: gap of unknown length
* 92166 94612: contig of 2447 bp in length
* 94613 98755: contig of 4043 bp in length
* 98756 98855: gap of unknown length
* 98856 102543: contig of 3688 bp in length
* 102544 102643: gap of unknown length
* 102644 105833: contig of 3190 bp in length
* 105834 105933: gap of unknown length
* 105934 108499: contig of 2566 bp in length
* 108500 108599: gap of unknown length
* 108600 111672: contig of 3073 bp in length
* 111673 111772: gap of unknown length
* 111773 114662: contig of 2890 bp in length
* 114663 114762: gap of unknown length
* 114763 116145: contig of 1383 bp in length
* 116146 116245: gap of unknown length
* 116246 118147: contig of 1902 bp in length
* 118148 118247: gap of unknown length
* 118248 120579: contig of 2332 bp in length
* 120580 120679: gap of unknown length
* 120680 122618: contig of 1939 bp in length
* 122619 122718: gap of unknown length
* 122719 124539: contig of 1821 bp in length
* 124540 124639: gap of unknown length
* 124640 126651: contig of 2012 bp in length
* 126652 126751: gap of unknown length
* 126752 128460: contig of 1709 bp in length
* 128461 128560: gap of unknown length
* 128561 130592: contig of 2032 bp in length
* 130593 130692: gap of unknown length
* 130693 132864: contig of 2172 bp in length
* 132865 132964: gap of unknown length
* 132965 134881: contig of 1917 bp in length
* 134882 134981: gap of unknown length
* 134982 136682: contig of 1701 bp in length
* 136683 136782: gap of unknown length
* 136783 139138: contig of 2356 bp in length
* 139139 139238: gap of unknown length
* 139239 140757: contig of 1519 bp in length
* 140758 140857: gap of unknown length
* 140858 142558: contig of 1701 bp in length
* 142559 142658: gap of unknown length
* 142659 144303: contig of 1645 bp in length
* 144304 144403: gap of unknown length
* 144404 145490: contig of 1087 bp in length
```

```
* 145491 145590: gap of unknown length
* 145591 147171: contig of 1581 bp in length
* 147172 147271: gap of unknown length
* 147272 148639: contig of 1368 bp in length
* 148640 148739: gap of unknown length
* 148740 150295: contig of 1556 bp in length
* 150296 150395: gap of unknown length
* 150396 151505: contig of 1110 bp in length
* 151506 151605: gap of unknown length
* 151606 152813: contig of 1208 bp in length
* 152814 152913: gap of unknown length
* 152914 155423: contig of 2510 bp in length
* 155424 155523: gap of unknown length
* 155524 156880: contig of 1357 bp in length
* 156881 156980: gap of unknown length
* 156981 158045: contig of 1065 bp in length
* 158046 158145: gap of unknown length
* 158146 159255: contig of 1110 bp in length
* 159256 159355: gap of unknown length
* 159356 160417: contig of 1062 bp in length
* 160418 160517: gap of unknown length
* 160518 161644: contig of 1127 bp in length
* 161645 161744: gap of unknown length
* 161745 163180: contig of 1436 bp in length
* 163181 163280: gap of unknown length
* 163281 164483: contig of 1203 bp in length
* 164484 164583: gap of unknown length
* 164584 166039: contig of 1456 bp in length
* 166040 166139: gap of unknown length
* 166140 167290: contig of 1151 bp in length

alignment_scores:
    Quality: 1059.00      Length: 835
    Ratio: 3.507         Gaps: 11
    Percent Similarity: 36.168    Percent Identity: 31.976

alignment_block:
US-09-357-675C-21 x AC105589 ..

Align seg 1/1, to: AC105589 from: 1 to: 178418

      2 LeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCysPr 18
      |||||
39095 CTGGGCTTCATCACCAGGCTCTCTCATCAACTCTGCTCTCTGTGTAC 39144

      18 OGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32
      |||
39145 CGGATACCGATTACCTCAAAATCTCAGTACTTTGTACTCAGCCAGGTAAC 39194

      32 .....
39195 ATATTTTATGTTCTTTGCTAGCACGCTAGTTGCTAAGTCCATGTTTATTA 39244

      32 .....
39245 CCGGAATAGTGGACCTGCCATCCACCAACATTAATGCGTGTACTCCAG 39294

      32 .....
39295 TAGCTTAGGGTTATTTAATGCTCTGGAAGATGAAACAAGTCGCTAGGG 39344

      32 .....
39345 TTGGTGGACACACCTCTAATCCACGACTCGGGAACAGGCAGGAGGAT 39394

      32 .....
39395 CAACTAGGAGTTTAAAGTATCATCCCGCTACATAGAAAGGTCTGCCTGGGC 39444

      32 .....
39445 TACAAAAAGACCCGGTCTCAAAGGGGAGGTAAATCAGAAAAAAAAG 39494
```

```
32 ..... 32
39495 TAAGTTCCGGCTCCACCCCACTTGAGCCTCCCAAGTTTCTCGGGCA 39544
33 .....ArgProArgAl 36
39545 CGCGGTAGAAATCCTGAGAAATCGCTGTGTCAATTCACAGGCCGACAGC 39594
36 aMetAlaIleSerSer...SerCysGluLeuProLeuValAlaValC 52
||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
39595 CATGGCAGAGTCCTCATCGACTTCCTGGGAGCTGCCCTGGTGTGTGT 39644
52 ysGlnValThrSerThrProAspLysGlnGlnAsnPhelYsThrCysAla 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39645 GCCAGGTAAACATCAACCAACCAAGCAAGCAAGAACTTTAAACGTGTGCT 39694
69 GluLeuValargGluAlaAlaAargLeuGlyAlaCysLeuAlaPheLeuPr 85
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39695 GAGCTGGTTCAAGAGGTACAGACTGGGGCTTGCCCTGGCCTTTCGTGCC 39744
85 oGluAlaPheAspPheIleAlaAargAspProAlaGluThrLeuHisLeuS 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39745 TGAGGCATTTGACTTATTACGACGAATCTCGCGAGACATTACTCCTGT 39794
102 erGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla... 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39795 CCGAACCACTGGATGGGACCTTTTGGGACAATATAGCCAGCTTGCCAGG 39844
117 ..... 117
39845 TACAAGGTAGGGTGGGGTGGGGGAGGGGTANAATCCCTGTGGGCAG 39894
117 ..... 117
39895 TGTTCCTGGATTGCCACATAGAGGGGTAGAGTTGTCAAGTGTCCCTCCC 39944
118 ..Arg.GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgG1 133
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39945 CTCAGGAATGTGAATCTGCCTGCCTTGGTGGTTTCCACCAACGTGG 39994
133 yGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuA 150
||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39995 CCAGAATGGGAACAGAATCCGAAATCATCCATTGTCTGGGCTTCTGA 40044
150 sn.SerLys..... 152
||: |||||
40045 ACACCAAGGTGAGACATTCCAATCCTTCAACTTGGCTTCTCTAAGGTC 40094
152 ..... 152
40095 ATCTAACTTCTCTCACTTAGAACTTTAAACATTTTCTCACTGCTTGCT 40144
152 ..... 152
40145 TTAGGGGTTTTTTGTTTTGTTTGAATGCTCTGTCTTAGCCATAAAA 40194
152 ..... 152
40195 TTACTCTTGGGAGAAGAGCTAAGCTGGACTTCAGTGTCTGTACTGAGC 40244
153 .....GlyAlaValAlaAlaThrTyrArgLysThrHisL 164
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40245 ATTTTCTCTACTTTAGGATCAGTAGTGGCCAGTTACAGGAAGACACATC 40294
164 euCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40295 TATGTGATGATAGATCCAGGTCAGGGCTATGAGAGAAAGCAACTAT 40344
181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLy 197
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40345 ACCATGCTGGATACGCTCTTGACCACTGTCAAGACACACAGGACAA 40394
197 sIleGly..Leu..... 200
||||| |||||
```



Percent Similarity: 74.150      Gaps: 3  
Percent Identity: 48.639

```

alignment_block:
  US-09-357-675C-21 x AF069989  ..
  Align seg 1/1 to: AF069989 from: 1 to: 1521

27 valLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSe 43
80 GTGATCGGCATTCATCAGCAGCTACCGAGATGCTCCGTCCAAAAGAGAA 129
43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
130 GATCAAAAGTCGCACCATTCCTCTGGTGCAGATCGCTCTACCGCGACA 179
60 ysgInGlnAsnPhelYsThrCysAlaGluLeuValArgGluAlaAlaArg 76
180 AGCGCGCTAATCTTAGCCAAGTCATACAGCTAGTGGATAGGCCCAAGTCA 229
77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
230 CAGAACGCTCGATGCTCTTCCTGCGTAGTGTGTGACTTTGTGGCGGA 279
93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuL 110
280 GAGCGCGGCCCAAACTATTGAGCTCTCCGAGGGCTTGGACGCGAGTTAA 329
110 euGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
330 TGGCGCAGTACCGGGAATTGGCGAAGTGCAACAAATTTGGATTTCCCTG 379
127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
380 GTGCGGTGCACGACGCGGAACGAT.....CAAAATATCTT 414
143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaIleValAla 160
415 CAACGCTCATGTTTCTCAACGAGAAAGGGAATAGCAGCAGTATACA 464
160 rgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPrometCys 176
465 GAAAGCTGCACATGTTTGATTGACTACGACTAAA...GAGGTTTCGCTACGC 511
177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
512 GAATCAGATACAGTTACGCGCGGATACTGCTTGAGCGCCCGCAGTGAGC 561
193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
562 TCCAGTTGGCCAGATAGGGCTTCAGATTTGCTACGACCTGCGTTTGTG 611
210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyPro 226
612 AGCCGGCGGTGCTCAGGAAGCTGGGTGCCAATTTGTTAACTATACCACA 661
227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
662 TCCGCATTACATACCACCGCTAGGCGCAGCTGGGAAATCCTCCTCGG 711
243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysG 260
712 GGCCAGAGCCATAGAGACTCAATGCTTTGTGGTGCCTGCGCTCGGCTCAGATAG 761
260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
762 GTTGGCACAAACAGACGACAGTTGGGCGCCACAGCATGATCGTTAGC 811
277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeuCy 292
812 CCCTGGGAAACGTACTGGCTGACTGCAGCGCAGCAGGAGCTTGATATAGG 861
292 sLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgHisLeuP 309
862 CACGGCCGAGGTGACCTTTCCGTTGTTCAATCTCTATCAGCCACCATGC 911

```





268	rTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC	285
911	TTATGCCACAGCATGGTGGTAGACCCCTGGGGAACAGTGGTGGCCGCT	960
285	ysSerGluGlyProGlyLeuCysLeuAlaAargIleAspLeuAsnTyrLeu	301
961	GCCTCTAGGGGGCAGGCCCTTCCTTGCCTTGCCTGGAATAGACCTCAACTATCTG	1010
302	ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeu	318
1011	CGACAGTGGCGGACACCTGCTGTGTTCACGACCGAGGCGCTGACCT	1060
318	uTyrGlyAsnLeuGlyHisProLeuSer	327
1061	CTATGCAATCTGGGTCAACCCACTGCT	1088
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: A2A		
seq_documentation_block:		
ID	AAF16257 standard; cDNA; 1382 BP.	
XX		
AC	AAF16257;	
XX		
DT	13-MAR-2001 (first entry)	
XX		
XX	Human prostate cancer antigen nucleotide sequence SEQ ID NO	
XX		
KW	Human; prostate cancer; prostate cancer antigen; detection;	
KW	neuroprotective; cytosolic; cardioactive; immunomodulatory	
KW	vulnary; gastrointestinal; nephrotropic; antineoplastic; g	
KW	antibacterial; gene therapy; neural; immune; reproductive; p	
KW	gastrointestinal; pulmonary; cardiovascular; proliferative	
KW	wound; infectious disease; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO2000055174-A1.	
XX		
PD	21-SEP-2000.	
XX		
PF	08-MAR-2000; 2000WO-US05988.	
XX		
PR	12-MAR-1999; 99US-0124270.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PA	(ROSE/) ROSEN C A.	
XX		
PI	Rosen CA, Ruben SM;	
XX		
DR	WPI: 2000-587513/55.	
DR	P-PSDB; AAB57054.	
XX		
PT	Prostate cancer associated gene sequences, referred to as p	
PT	cancer antigens, useful for treatment, prevention, and diag	
PT	disorders such as prostate cancer .	
XX		
XX		
PS	Claim 1; Page 1124; 2338pp; English.	
XX		
CC	AAF15566 to AAF16505 encode the human prostate cancer associ	
CC	proteins, called prostate cancer antigens, given in AAB5636	
CC	The prostate cancer antigens can have neuroprotective, cyto	
CC	cardioactive, immunomodulatory, muscular, vulnary, gastro	
CC	nephrotropic, antiinfective, gynaecological and antibacteri	
CC	can be used in gene therapy. The prostate cancer antigen	
CC	polynucleotides may be used for detection of prostate cancer	
CC	identification, as chromosome markers, and for numerous oth	
CC	or research purposes. The prostate cancer antigens may be u	
CC	disorders such as neural, immune, muscular, reproductive,	
CC	gastrointestinal, pulmonary, cardiovascular, renal, and pro	
CC	disorders, wounds, and infectious diseases. AAF16506 to AAF	
CC	AAB57303 represent sequences used in the exemplification of	
XX		
XX		



SQ Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

alignment\_scores:                      Length:        327  
Quality: 1718.00  
Ratio: 5.270                      Gaps: 0  
Percent Similarity: 99.694        Percent Identity: 99.694

alignment\_block:  
US-09-357-675C-21 x AAF16257 ..

Align seg 1/1 to: AAF16257 from: 1 to: 1382

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17  
77 ATGCTGGGCTTCATCACCAGGCTCCCTCACAGATTCTCTGCTCCCTTCGTG 126

17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgp 34  
127 TCCTGGACTCCGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGC 176

34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50  
177 CCAGAGCATGGCTATCTCTCTCTCTCTCCCTGGCACTGCCCTGGTGGCT 226

51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67  
227 GTGTGCCAGGTACATCGAGCCAGCAAGCAAGCAAGCAAGCAAGCAAG 276

67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
277 TGTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGGCTGGCTGGCTTTC 326

84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
327 TGCCTGAGGCAATTGACTTCATTGACGGGACCTTCGAGAGCCCTACAC 376

101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluThrThrGlnLeuAl 117  
377 CTGTCTGAACCACTGGTGGGAAACTTTTGGAGAATACACCCAGCTTGC 426

117 aArgGluCysGlyLeuThrLeuSerLeuGlyGlyPheHisGluArgGlyG 134  
427 CAGGGAATGTGGACTCTGGCTGTCTCTGGTGGTGGTGGTGGTGGTGG 476

134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150  
477 AAGACTGGGAGCAGACTCAGAAAATCTACAAATGTCTACGTCGTCTGAA 526

151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167  
527 AGCAAGGGGCGAGTAGTGGCCACTTCAGGAAGACACATCTGTGTGACGT 576

167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
577 AGAGATTCCAGGCGAGGGGCT.ATGTGTGAAAGCAACTCTACCATGGCTG 625

184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
626 GGCCAGCTCTTGACTCACCCTGTGACACACAGCAGGCAAGATTGGTCTA 675

201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217  
676 GCTGTCTGCTATGACATCGGCTTCCTCGAACTCTCTCTGGCATTTGGCTCA 725

217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234  
726 AGCTGGAGCAGAGATACTTACCTATCTTACGCTTTTGGATCCATTACAG 775

234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250  
776 GCCCAGCCCACTGGGAGGTGTGTGCGGGGCCCTGTCTATCGAAACCCAG 825

251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267  
826 TGCTATAGTGGCAGCAGCACAGTGTGGACGCCACCATGAGAAGAGAGC 875

267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaA 284  
876 AAGTTATGGCCACAGCATGTTGGTAGACCCCTGGGGAACAGTGGTGGCC 925

284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300  
926 GCTGCTCTGAGGGCCAGGCTCTGCTTGGCCGAATAGACCTCAACTAT 975

301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317  
976 CTGCGACAGTTGGCCGACACACCTGCTGCTTCCAGCACCGCAGGCCTGA 1025

317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327  
1026 CCTCTATGGCAATCTGGGTCAACCACTGTCT 1056

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AA30398

seq\_documentation\_block:  
ID AAX30398 standard; DNA; 1203 BP.

XX  
AC AAX30398;  
XX  
DT 14-MAY-1999 (first entry)  
XX  
DE DNA encoding a human secreted protein.

XX  
KW Secreted protein; cancer; tumour; neurodegenerative disorder;  
KW developmental abnormality; foetal deficiency; blood disorder;  
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;  
KW renal disease; diabetes; inflammation; allergy; ischemic shock;  
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;  
KW prostate disease; asthma; osteoporosis; arthritis; ss.

XX  
OS Homo sapiens.  
XX  
PN WO9907891-Al.  
XX  
PD 18-FEB-1999.  
XX  
PF 04-AUG-1998; 98WO-US16235.  
XX  
PR 19-AUG-1997; 97US-0056732.  
PR 05-AUG-1997; 97US-0054798.  
PR 05-AUG-1997; 97US-0054803.  
PR 05-AUG-1997; 97US-0054804.  
PR 05-AUG-1997; 97US-0054806.  
PR 05-AUG-1997; 97US-0054807.  
PR 05-AUG-1997; 97US-0054808.  
PR 05-AUG-1997; 97US-0054809.  
PR 05-AUG-1997; 97US-0055309.  
PR 05-AUG-1997; 97US-0055310.  
PR 05-AUG-1997; 97US-0055312.  
PR 05-AUG-1997; 97US-0055386.  
PR 05-AUG-1997; 97US-0055311.  
PR 18-AUG-1997; 97US-0055970.  
PR 18-AUG-1997; 97US-0055986.  
PR 19-AUG-1997; 97US-0056365.  
PR 19-AUG-1997; 97US-0056366.  
PR 19-AUG-1997; 97US-0056557.  
PR 19-AUG-1997; 97US-0056370.  
PR 19-AUG-1997; 97US-0056371.  
PR 19-AUG-1997; 97US-0056563.  
PR 19-AUG-1997; 97US-0056731.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, NI J;  
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;

XX WPI; 1999-167452/14.  
DR P-PSDB; AAY10877.  
XX  
PT New isolated human genes encoding secreted polypeptides - useful for  
PT diagnosis and treatment of pathological diseases  
XX  
PS Claim 3; Page 265-266; 331pp; English.  
XX  
CC The specification describes secreted proteins and their corresponding  
CC polynucleotides which are useful for preventing, treating or ameliorating  
CC medical conditions, e.g. by protein or gene therapy. Pathological  
CC conditions can also be diagnosed by determining the amount of the  
CC secreted polypeptides in a sample or by determining the presence of  
CC mutations in the polynucleotides. Specific uses are described for each  
CC of the products, based on which tissues they are most highly  
CC expressed in, and include developing products for the diagnosis or  
CC treatment of cancer, tumours, neurodegenerative disorders, developmental  
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,  
CC diseases of the immune system, autoimmune diseases, hepatic and renal  
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's  
CC and cognitive disorders, schizophrenia, cardiovascular disorders,  
CC prostate diseases, asthma, disorders involving osteoclasts such as  
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or  
CC thymus, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
XX  
SQ Sequence 1203 BP; 292 A; 307 C; 325 G; 278 T; 1 other;

## alignment\_scores:

Quality: 1067.00 Length: 207  
Ratio: 5.205 Gaps: 0  
Percent Similarity: 99.034 Percent Identity: 99.034

## alignment\_block:

US-09-357-675C-21 x AAX30398

Align seq 1/1 to: AAX30398 from: 1 to: 1203

121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyClnAspTrpG1 137  
|||||  
3 GGACTCGGCTGCTGGGTGGTGGTTCATGAGCGTGGCCAAAGACTGGGA 52  
137 uGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyA 154  
|||||  
53 GCAGACTCAGAAATCTACAAATGTCTACGTGCTGCTGAACAG. AAAGGG 101  
154 laValValAlaThrTyrArgLysThrHisLeuCysAspValGluIlePro 170  
|||||  
102 CAGTAGTGGCCACTTACAGGAAGACACATCTGTGACGTAGAGATTCCA 151  
171 GlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyProSerLe 187  
|||||  
152 GGCAGGGGCT. ATGTGTGAAGCAACTCTACCATGCTGGGCCAGTCT 200  
187 uGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCyst 204  
|||||  
201 TGAGTCACCTGTCAGCACACCACAGCAGCAAGATTGGTCTAGTGTCTGCT 250  
204 yrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220  
251 ATGACATGCGGTTCCCTGAACCTCTCTGGCAATTGGCTCAAGCTGAGCA 300  
221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHi 237  
301 GAGATACTTACCTATCTTACGCTTTTGGATCCATTACAGGCCACGCCCA 350  
237 sTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCystTyrVal 254  
351 CTGGGAGGTGTCTGCGGGCCGCTGCTATCGAAACCCAGTGTATGTAG 400  
254 alAlaAlaLaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGly 270

|||||  
401 TGGCAGCAGCACAGTGTGGAGCCACCACCATGAGAAGAGAGCAAGTTATGGC 450  
271 HisSerMetValValAspProTrpGlyThrValValAlaAArgCysSerG1 287  
|||||  
451 CACAGCATGTGTAGTACCCCTGGGAACAGTGGTGGCCGCTGCTCTGA 500  
287 uGlyProGlyLeuCysLeuAlaAArgIleAspLeuAsnTyrLeuArgGlnL 304  
|||||  
501 GGGCCAGGCTCTGCTTGGCCGATAGACCTCAACTATCTGGACAGT 550  
304 euArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGly 320  
|||||  
551 TGGCCGACACCTGCTGTGTTCCAGCACCGCAGGCTGACCTCTATGGC 600  
321 AsnLeuGlyHisProLeuSer 327  
601 AATCTGGGTCAACCCACTGTCT 621  
seq\_name: /SIDS1/gcgdata/hold- geneseq/geneseqn-emb1/NA2002.DAT.AAD25458  
seq\_documentation\_block:  
ID AAD25458 standard; cDNA: 1214 BP.  
XX AC AAD25458;  
XX DT 26-MAR-2002 (first entry)  
XX DE Xenopus laevis Nit1 cDNA.  
XX KW Frog; Nit1; cytostatic; neuroprotective; cellular pathway; therapy;  
XX KW apoptosis; proliferative disorder; degenerative disease; ss.  
XX OS Xenopus laevis.  
XX FH Key Location/Qualifiers  
FT CDS 180..1046  
FT /\*tag= a  
FT /\*product= "Frog Nit1 protein"  
XX WO200187958-A2.  
PN 22-NOV-2001.  
XX PD 15-MAY-2001; 2001WO-US15664.  
XX PF 16-MAY-2000; 2000US-204713P.  
XX PR (UYJE-) UNIV JEFFERSON THOMAS.  
XX PA Croce C, Brenner C, Pekarski Y;  
XX PI WPI; 2002-082984/11.  
XX DR P-PSDB; AAE15792.  
XX PT Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,  
XX PT useful to find molecules that mimic or antagonize Fhit interaction for  
XX PT the treatment of proliferative or degenerative diseases  
PS Claim 21; Page 60; 61pp; English.  
XX The invention relates to isolated nucleic acids comprising a fully  
XX defined cDNA nucleotide sequence encoding human, Xenopus laevis and  
XX mouse Nit2 proteins. Nit and Fhit proteins are encoded as fusion  
XX proteins in invertebrates and as separate polypeptides in vertebrates.  
XX Nit and Fhit interact physically and functions in same cellular pathways.  
XX Molecules which bind Nit2 and mimic or antagonise Fhit interaction are  
XX used to treat diseases in which activity of Nit2 protein is altered in  
XX a mammal. Fhit mimics induce apoptosis and are particularly useful to  
XX treat proliferative disorders whilst Fhit antagonists promote cell  
XX proliferation and are particularly useful to treat degenerative disease.  
XX The present sequence is frog Nit1 cDNA.

SQ Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

alignment\_scores:  
Quality: 968.50 Length: 307  
Ratio: 3.905 Gaps: 1  
Percent Similarity: 80.782 Percent Identity: 57.980

alignment\_block:

US-09-357-675C-21 x AAD25458 ..

Align seg 1/1 to: AAD25458 from: 1 to: 1214

16 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32  
||||| : : : : :  
144 CTGTGTCGGATATAGGGGTAGAGCCCTGAGTCCAAATG..... 182  
32 oArgProArgAlaMetAlaIleSerSerSerCysGluLeuProLeuV 49  
: : : : :  
183 .....GCTGGTGCCACAAAGCCCTGA 204  
49 alaIaValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelys 65  
: : : : :  
205 TTGCGGTGCGGATGACCTTCAACCTCTGATAAGGAGAAATTTCCGCC 254  
66 ThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAl 82  
||||| : : : : :  
255 ACGTGTTCCGGCTGATCCGGAGGCTGCGGGCGTCCGCTTGCAATGGT 304  
82 aPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrL 99  
: : : : :  
305 GTTTCTGCCGAAGCCTTTGACTATATATCGGGGCGAGCATTGAGGAGACGC 354  
99 euHisLeuSerGluProLeuGlyClyLysLeuLeuGluGluThrGln 115  
||| : : : : :  
355 TGAGTCTGGCTGAGTCTCTCATGTGGGACACCATTTACGCGTTACACCCAA 404  
116 LeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyPheHisGluAr 132  
||||| : : : : :  
405 CTCGCCAGGAGGTGGGCTGCTGCTTCCCTGGGGGATTTCAATGAGAA 454  
132 gGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuL 149  
: : : : :  
455 AGGACCAACTGGGACACGCGACCAACATTTCCAATTTCTCACGTGTTG 504  
149 euAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCys 165  
: : : : :  
505 TGGACAACACAGGCGACATAGTATCGGTGTACCGCAAGGCTCACCTGTT 554  
166 AspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMe 182  
||||| : : : : :  
555 GACGTACACTGCAGATGGAGTGTCTACTCAGAGAGAGAGTCCACCT 604  
182 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 199  
||||| : : : : :  
605 CCCGAGACAGAGCTTATTCGCCCCATCACTTCTCCACGAGGAAGATTG 654  
199 lyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeu 215  
||||| : : : : :  
655 GCTTGGGGGTGTATGACGCTCCGCTCCAGAAATTCCTCTGCGCTGTG 704  
216 AlaGlnAlaGluAlaGluIleLeuThrTyrProSerAlaPheGlySerIl 232  
||||| : : : : :  
705 GCCCAACAGGAGCAGAGACTTCTACATTACCTTCTGCTTCCCTTCCAC 754  
232 eThrGlyProAlaHisTrpGluValLeuLeuArgAlaAlaIleGluT 249  
||||| : : : : :  
755 TACTGGTCTGGCACATTTGGGAGGTGTGCTCAGAGCCCGTGCATAGAAA 804  
249 hrGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisGluLys 265  
||||| : : : : :  
805 CCCAGTCTAGTATGTGACGCGGCACAGACAGACACATGATGAAG 854

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL12225

seq\_documentation\_block:

ID ABL12225 standard; cDNA; 1495 BP.

XX ABL12225;

AC ABL12225;

XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers BW;

XX DR WPI; 2001-656860/75.

XX DR P-PSDB; ABB68122.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Claim 1; SEQ ID NO 31157; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;

alignment\_scores:

Quality: 710.50 Length: 294

Ratio: 3.259 Gaps: 3

Percent Similarity: 74.150 Percent Identity: 48.639

## alignment\_block:

US-09-357-675C-21 x ABL12225

Align seg 1/1 to: ABL12225 from: 1 to: 1495

```
27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSe 43
|||||:  ::  ::  ||| |||:|||||:  ::
80 GTGATGCCATTTCATCAGCAGCTACGACGATGTCGCTCCAAAGAGAA 129
|||:|||||:  ::  ::  ||| |||:|||||:  ::
43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
130 GGATCAAAAGTCGACCATTCCTGCTGGCTCAGATCGCTCACCAGCGACA 179
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
60 ySGlnGlnAsnPheLeuValThrCysAlaGlnLeuValArgGluAlaAlaArg 76
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
180 AGCGGGCTAATCTTACCAAGTATAGAGTAGTGGATAGGGCCAAAGTCA 229
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
230 CAGAACGCTGCTATCTTCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTG 279
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuL 110
:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
280 GAGCCGGACCCAACTATTGAGCTCCGAGGCTTGGACGGCGAGTTAA 329
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
110 euGluGluTyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
330 TGGCGCAGTACCGGGAATTCGCAAGTGCACAAATTTGGATTTCCCTG 379
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
127 GlyGlyPheHisGlnArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
380 GGTGGCTGCAGCAGCGGAACGAT.....CAAAAAATCTT 414
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaThrTyra 160
:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
415 CAACGCTCATGTTTGTCTCAACGAGAAGGGAACTAGCAGCATATACA 464
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
160 rGlyThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
465 GAAAGCTGCACATGTTGATGTACGACTAAA...GAGGTTCCGCTACGC 511
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
512 GAATCAGATACATGTCGCGGGATAGCTGCTTTCAGCGCCAGTGCAGCAC 561
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
562 TCCAGTTGGCCAGATAGGCTTCAGATTTGCTACGACCTCGCTTTGCTG 611
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyro 226
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
612 AGCGGGGTGCTGCTCAGGAAGCTGGTGCCAAATTTGTTAACTACCCA 661
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
227 SerAlaPheGlySerLeuThrGlyProAlaHisTrpGluValLeuLeuAr 243
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
662 TCCGCATTCACATACGACCGCTAAGCGGCACCTGGGAATCTCTCTGCG 711
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
243 gAlaArgAlaIleGluThrGlnCysTyrrValValAlaAlaGlnCysG 260
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
712 GGCAGAGCATAGAGACTCAATGCTTGTGGTGGCTGGCGCTCAGATAG 761
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
260 lyArgHisHisGluLysArgAlaSerTyrrGlyHisSerMetValValAsp 276
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
762 GTTGGCACACACAGAGCGACAGTGTGGGCCACAGCATGCTCGTTAGC 811
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
277 ProTrpGlyThrValValAlaArgCysSerGluGlyPro...GlyLeuCy 292
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
812 CCCTGGGGAACACGTACTGGCTGACTCAGCGCAGGAGCTTCATATAGG 861
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
292 sLeuAlaArgIleAspLeuAsnTyrrLeuArgGlnLeuArgArgHisLeuP 309
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
862 CACGGCCGAGGTGGACTTTCGCTGCTTCAATCTCTGTATCAGACCATGC 911
|||:|||||:  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
```

```
309 roValPheGlnHisArgArgProAspLeuTyrr 319
|||:|||||:  ::  ::  ||| |||:|||||:  ::
912 CCTGCTCGAACATCGTCGAACGACATCTAC 943
```

seq\_name: /SIDSL/cgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL12224

## seq\_documentation\_block:

ID ABL12224 standard; cDNA; 3548 BP.

AC ABL12224;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-658860/75.

XX P-PSDB; ABB68121.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

XX Claim 1; SEQ ID NO 31154; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;

## alignment\_scores:

Quality:	710.50	Length:	294
Ratio:	3.259	Gaps:	3
Percent Similarity:	74.150	Percent Identity:	48.639

## alignment\_block:

US-09-357-675C-21 x ABL12224

Align seg 1/1 to: ABL12224 from: 1 to: 3548

27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSe 43

1133 GTGATGCCATTTCATCAGCAGCTACGACGATGTCGCTCCAAAGAGAA 1182

43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60

```
1183 GGATCAAGAGTGGACCACTTCTGGCTCAGATCGCTCTACCAGCGACA 1232
60 ysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArg 76
1233 AGCGGGCTAATCTTAGCAAGTAGTATGAGTAGTGGATAGGGCCAACTCA 1282
77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
1283 CAGACGCTGCATGCTCTTCTCCCTGAGTGTGACTTTGTGGCGGA 1332
93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuL 110
1333 GAGCGGACCCAACTATTGAGCTCTCCGAGGCTTGGACGGCAGTTAA 1382
110 euGluGluTyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
1383 TGGCGCAGTACCGGAATTCGCGAAGGACCAAAATTTGGATTCCTCG 1432
127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
1433 GTGGCGTCGACGAGCGGAACGAT.....CAAAAAATCTT 1467
143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrA 160
1468 CAACGCTCATGTTTGTCTAACGAGAAGGGAACCTAGCAGCAGTATACA 1517
160 rGlyThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
1518 GAAGCTGCACATGTTGATGTTACGACTAAA...GAGGTTGCGCTACGC 1564
177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
1565 GAATCAGATACAGTACGCGCGGATACGCTTGAGCGCCGACGTGACAC 1614
193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPhePro 210
1615 TCCAGTTGGCCAGATAGGCTTCAGATTTGCTACGACCTGCTTTGCTG 1664
210 luLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
1665 AGCGCGCGGTGCTGCTCAGGAGCTGGTGCCAAATTTGTTAAACATACCA 1714
227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
1715 TCGCATTCACATACGCAACCGGTAAAGCGCAGCTGGGAATCTCTCTGG 1764
243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnCysG 260
1765 GGCAGAGCCATAGAGACTCAATGCTTGTGCTGCTGCGCTCAGATAG 1814
260 lYArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
1815 GTTGGCAACACCGACGACGACAGTGGGGCCACAGCATGCTGTTAGC 1864
277 ProTrpGlyThrValValAlaAlaArgCysSerGluGlyPro...GlyLeuCy 292
1865 CCCTGGGGAACGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1914
292 sLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgHisLeuP 309
1915 CACGGCGAGGTGGACCTTTCGCTGCTTCAATCTCTATCAGACCATGC 1964
309 roValPheGlnHisArgArgProAspLeuTyr 319
1965 CCTGCTTCGACATCGTCGAACACATCTAC 1996
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001B.DAT:ABL20222

seq\_documentation\_block:

ID: ABL20222 standard; DNA; 5692 BP.

XX  
AC

```
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEXE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions
XX Claim 1; SEQ ID NO 12139; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(CC (ABB57737-ABB72072)).
XX The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from wipo
at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;

alignment_scores:
Quality: 710.50 Length: 294
Ratio: 3.259 Gaps: 3
Percent Similarity: 74.150 Percent Identity: 48.639

alignment_block:
US-09-357-675C-21 x ABL20222 ..
Align seg 1/1 to: ABL20222 from: 1 to: 5692
27 ValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerSe 43
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2792 GTGATGCCCATTCATCAGCAGCTACGAGGATGTCCTCCCAAGAGGAA 2841
43 rCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspL 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2842 GGATCAAGTCCGACCATTCGCTGGGCTCAGATCGCTCTACCAGCGACA 2891
60 ysGlnGlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArg 76
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2892 AGCGGGCTAATCTTAGCAAGTAGTATGAGTAGTGGATAGGGCCAACTCA 2941
77 LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaAr 93
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2942 CAGACGCTGCATGCTCTTCTCCCTGAGTGTGACTTTGTGGCGGA 2991
93 gAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLeuL 110
```





[illegible][illegible]



```

|||||  ...  |||||||  |||||||  |||||||  |||||||
409 GGGAAATACGTTTCAAGAAATCAAAACATGTAGCCCTGGTGATAGTTT 458
187 uGluSerProValSerThrProAlaGlyLysIleGlyLeuAlaValCyst 204
459 C...TCCACATTTGATACGCTTACTGCAAGTGGGCTGGGCATCGCT 505
204 yRAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220
506 ATGATATGCGCTTCGCGAGCTTGCACAAATCTATGCACAAAGAGGCTGC 555
221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaH1 237
556 CAGCTCTGTGTATCTCTGGAGCTTCAATCTGACCACAGCAGCAGCCCA 605
237 sTrpGluValLeuLeuAlaArgAlaIleGluThrGlnCysTyrValV 254
606 CTGGAGCTGCTTCAGCGAGCCCGGCTGTGATAATCAGGTGTATGTGG 655
254 aLaLaAlaGlnCysGlyArgHisGluLysArgAlaSerTyr... 269
656 CTACAGCTCTCTCTGCT.....CGGATGACAAAGCTCGTATGTG 696
270 .....GlyHisSerMetValValAspProTrpGlyThrValValAlaAr 284
697 GCCTGGGGACACAGCACTGTGTGTGATCTTGGGGCAGGTCCTTAACCA 746
284 qCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrL 301
747 AGCTGGGCACGGAGGAACAATCTGTACTCAGACATAGACTGAAGAAGC 796
301 euArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 317
797 TGGCTGAATTCGCGACCAATCCCATTTTAAACACAGAACAGCAGAC 846
318 LeuTyr 319
|||||
847 CTCTAT 852

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.AAA81479
seq_documentation_block:
ID AAA81479 standard; DNA; 69936 BP.
XX
AC AAA81479;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
XX
PR 30-APR-1999; 99US-0132068.
XX
PR (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
XX WPI; 2000-318079/27.
XX
XX

```

```

PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
XX other Neisserial infections, for example, N.gonorrhoea
XX
XX Claim 7: Page 547-567; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
XX other more variable regions.
XX
SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

```

```

alignment_scores:
Quality: 404.00 Length: 307
Ratio: 2.138 Gaps: 8
Percent Similarity: 61.564 Percent Identity: 35.179

alignment_block:
US-09-357-675C-21 x AAA81479/rev ..
Align seg 1/1 to reverse of: AAA81479 from: 1 to: 69936

16 LeuCysProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPr 32
|||||
27227 TTGTGTAAA.....CGGTTTTCGAAAGTACCGTCATGCTAGCCTATCG 27184

32 oArg.....ProArgAlaMetAlaIleSerSerSerSerCysGluLeuP 47
|||||
27183 CGGGAATATGCCGAGGTGCGCAGGAGAAAGAGAGAGAAATGGACANAA 27134

47 roLeuValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsn 63
|||||
27133 TCAGAGTTGCGCGCTCAGATGCTGCGCGGTGTCGCCGGAACCAAC 27084

64 PheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCy 80
|||
27083 GTCGCCGCCATGAAACGCTGGTGCACGCGCGCGGAGGAGGTGCGGA 27034

80 sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 97
|||||
27033 TTGGTGCTGCTGCCCGCAATATTGGGTGCTGATGGCGCAACGATACCG 26984

97 luThrLeuHisLeuSerGluProLeu...GlyGlyLysLeuLeuGluGlu 112
|||
26983 ACAAACTCGCGCTTCGCGAGCCITTTGGCGCGGAGCCGCTTTCACACGGCA 26934

113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
|||||
26933 TTGAGCGAAACGGCGAAAGAAATGCGGCTGTGTGCTG...TTCGCGGGAC 26887

```





```
64  PheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaC 80
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454104 GTGCGCGCCATGAACAGCTGTGTCGACGGCGGCGAGCGGTGGGA 454153
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
80  sLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaG 97
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454154 TTGGGTGCTGCTCCGCAATATTGGTGTGTGATGGCGCAACGATACCG 454203
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
97  LuThrLeuHisLeuSerGluProLeu...GlyClyLysLeuLeuGluGlu 112
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454204 ACAACTCGCGCTGGCGAGCCCTTGGGCGGCGGACGCTTCAGACGGCA 454253
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
113 TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGly.. 128
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454254 TTGAGCGAAGCGGCAAGAAAGATCGCGGTGCTG...TTCGGCGGGAC 454300
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
129 .....PheHisGluArgGlyGlnAspTrpGluGlnThrGlnL 141
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454301 TGTGCGCGCTGCAAGCTGCGAGCGGGT.....A 454329
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
141 yIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454330 AAGTGATGAATACGCTGTGGTGTACGGCGGACGGCGTAAGACGGGG 454379
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454380 CTGTACCACAAATGCACCTC.....TTCGGTTCGCGTTGGCGGA 454423
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454424 ACCTATGCGGAGCGGATACCATCGCGCGCGGGGATGTGCGCACT 454473
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
191 alSerThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArg 207
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454474 TGTGCGGAGAGCGTGGCGGTGGCGGCGGCAATGTTACGATGTCCG 454523
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
208 PheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuTh 224
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454524 TTTCCCGAA.....TTTTCGACGCCAGTTGCGCTTTCAGCTATGAT 454567
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
224 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValL 241
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454568 GCTGCCCGCTGCGTTCACGCACACGCGGCAAGCGCATTTGGGAGCTGC 454617
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
241 euLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAla 257
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454618 TGTGCGCGCGCGTGGCGTGAACCAATGTTACGTGCGCGCGGCA 454667
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
258 GlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetVa 274
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454668 CAGGCGCGTTTGACGAAACGCGGCGCACCGTTCGGACACACATGAT 454717
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
274 lValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGlyL 291
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454718 TGTGCGATCGCTGGGCGGCGTGTTCAGCTATTGCGCGGAGGCGGAG 454767
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
291 euCysLeuAlaArgIleAspLeuAspTyrLeuArgGlnLeuArgArgHis 307
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454768 TTGTTACGCAGACATCGATGCCAACCGCTGAACAGCGTCCGCAACCG 454817
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
308 LeuProValPheGlnHisArg 314
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
454818 CTGCCCGCTTGAATACCGG 454838
```

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT: AAC77176

seq\_documentation\_block:

ID AAC77176 standard; cDNA; 1725 BP.

XX AAC77176;

XX DT 08-FEB-2001 (first entry)

XX

Human OREF ORF2731 polynucleotide sequence SEQ ID NO:5461.

Human: open reading frame; OREF; detection; cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

W0200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB42967.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

Claim 5; Page 4641-4643; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human OREF open reading frames 1 to 3161. The OREF sequences have activities such as: cytostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; antirheumatic; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an OREF-associated disorder. The nucleic acids can be used to express OREF proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 1725 BP; 466 A; 368 C; 372 G; 517 T; 2 other;

alignment\_scores:

Quality: 392.00

Ratio: 1.858

Percent Similarity: 43.867

Length: 481

Gaps: 14

Percent Identity: 24.324

## alignment\_block:

US-09-357-675C-21 x AAC77176

Align seg 1/1 to: AAC77176 from: 1 to: 1725

```
23 ProGlnLeuSerValLeuCysAlaGlnProArgProHrgAlaMetAla11 39
|||||: : : : :
11 CCAGAGTGGTGTGCTGC.....AGAGTCATG..... 40

39 eSerSerSerCysGlnLeuProLeuValAlaValCysGlnValThrS 56
: : : : : : : : : : : : : : : : : : : : : : : : : :
41 .....ACCTCTTCGCGTGGCCCTCATCCAGCTT...CAGATTTCCT 80
||
56 erThrProAspLysGlnGlnAsnPhelYsThrCysAlaGluLeuValArg 72
||
81 CCATC.....AAATCAGATAAGCTCACTCGCGCTTGTAGCTTCATCCGG 124

73 GluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAs 89
|||||: : : : : : : : : : : : : : : : : : : : : : :
125 GAGCAGCAACGCAAGGAGCCAAATAGTTTCTTGC CGGAATGCTTT.. 172

89 pPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuG 106
: : : : : : : : : : : : : : : : : : : : : : : : : :
173 .....AATTCCTCCATAG 185

106 lYgLYsLeuLeuGluGluTyr..... 113
||: : : : : : : : : : : : : : : : : : : : : : : :
186 GAGGAAATATTTTCCTGAATATGCAGAGAAATTCCTGGTGAATCCACA 235

114 .....ThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG1 127
: : : : : : : : : : : : : : : : : : : : : : : : : :
236 CAGAAGCTTTCTGAAGTAGCAAGAAATAGTATATATCTCATTTGGAGG 285

127 yGlyPhe.HisGluArgGlyGln...AspTrpGluGlnThrGlnLysIle 142
||: : : : : : : : : : : : : : : : : : : : : : : :
286 TAACCTCTACCCACAGGCTCTATCCCTGAAGAGGATGCTGGAAATTA 335

143 TyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTy 159
|||||: : : : : : : : : : : : : : : : : : : : : : :
336 TATAACACCTGTGTGTGTTGGCCCTGATGGAACCTTTACTAGCAAAAGTA 385

159 rArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetC 176
|||||: : : : : : : : : : : : : : : : : : : : : : :
386 TAGAAAGATCCATCTGTTTGCATATGATGTTCTCTGGAAATATACATTTC 435

176 ysGluSerAsnSerThrMetProGlyProSerLeuGlu..... 188
|||||: : : : : : : : : : : : : : : : : : : : : : :
436 AAGAATCTAAACATTCAGTCCGGGTGATAGTTTCTCCACATTTGATACT 485

188 ..... 188

486 CGTATGTACCAGATAAGTTTGCCTCTTTAGCAATCTCAGTAGAGACAAT 535

189 .....SerProValSer..... 192
|||||: : : : : : : : : : : : : : : : : : : : : : :
536 CAGGTATTATTCTTTTGTCTCTCTCCGATTCTTCACATAACCTAA 585

192 ..... 192

586 CTGAAGACCATAACTGAGAAAGCAGAGAAATCATCAGATCTGGAAG 635

192 ..... 192

636 TTCGGCTTATTTCAGAACTAAGGATTGACACGATTTTGCCTTTGATT 685

192 ..... 192

686 TGATTGTAGCTTCTGTTACGGCTTCCAGAGATATACCTATTAGGCTACAG 735

192 ..... 192

736 TTGAGTACCTCCCATCTAGATAATAAGCATTTCAATTAGAATGATTTCTC 785
```

```
193 .....ThrProAla.GlyLys..... 197
|||||: : : : :
786 ATCTTTACTCCGCTGATGTAATGATGCTCTTTATGAGATGAAGTCCAAGT 835

197 ..... 197

836 AGGAATGAGCTTGTAAATTATCTCTGTCTCAGGTCCTGTGTTAAATTTAT 885

197 ..... 197

886 CCCTGTCAGTGTTTGTGATCATTATGTCATGAGGATTTCCCTCCCTGCCAC 935

197 ..... 197

936 ACCATGCTGTAGGAGTAACTTTTCATTTCGTCATTTCTGTGTTGGAAA 985

198 .....IleGlyLeuAlaValCysTyrAspMetArgPhePro 209
: : : : : : : : : : : : : : : : : : : : : : : :
986 CAGCTTACTGCAGAGTGGTCTGGGCATCTGCTACGACATCGCGTTTGCA 1035

210 GluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPr 226
|||||: : : : : : : : : : : : : : : : : : : : : : :
1036 GAGCTTGCACAAATCTACGCACAGAGGCTGCCAGCTGTGTGATATCC 1085

226 oSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuA 243
||: : : : : : : : : : : : : : : : : : : : : : : :
1086 AGGAGCTTTTAATCTGACCACTGACCCAGCCCATTTGGAGTTACTTCAGC 1135

243 rgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnCys 259
|||||: : : : : : : : : : : : : : : : : : : : : : :
1136 GAAGCCGGCTGTGATAATCAGGTGTATGTGCCACAGGCTCTCTCCGCC 1185

260 GlyArgHisGlyLysArgAlaSerTyr.....GlyHisSerme 273
||: : : : : : : : : : : : : : : : : : : : : : : :
1186 .....CGGATGACAAAGCCTCTCTATGTCCTGGGACACAGCAC 1226

273 tValValAspProTrpGlyThrValValAlaArgCysSerGluGlyProG 290
|||||: : : : : : : : : : : : : : : : : : : : : : :
1227 CGTGTGAACCTTGGGGGAGGTTCTAGCCAAAGCTGGCAGACAGAAGA 1276

290 lYeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArg 306
: : : : : : : : : : : : : : : : : : : : : : : :
1277 CAATCGTGTATTCAGACATACCTGAAGAAGCTGGCTGAAATACGCCAG 1326

307 HisLeuProValPheGlnHisArgArgProAspLeuTyr 319
: : : : : : : : : : : : : : : : : : : : : : : :
1327 CAATCCCGCTTTTAGACAGAGCGATCAGACCTCTAT 1365
```

seq\_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABLL13787

## seq\_documentation\_block:

ID: ABL13787 standard; cDNA; 898 BP.

XX ABL13787;

XX AC ABL13787;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX XX

PA (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR P-PSDB; ABB69684.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1; SEQ ID NO 35843; 2lpp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB163051), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 898 BP; 243 A; 236 C; 231 G; 188 T; 0 other;

alignment\_scores:  
 Quality: 387.00 Length: 293  
 Ratio: 2.138 Gaps: 5  
 Percent Similarity: 61.775 Percent Identity: 31.399

alignment\_block:  
 US-09-357-675C-21 x ABL13787 ..  
 Align seg 1/1 to: ABL13787 from: 1 to: 898

```

41 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerTh 57
   |||:||||:||||:||||:||||:||||:||||:||||:||||:
50 AGCAAACTAGCACATCATCGTGGTGGCGCTGCTGCAGCTCAAGGGTTC 99
   |||:||||:||||:||||:||||:||||:||||:||||:
57 rProAspLysGlnGlnAsnPhelysthrCysAlaGluLeuValArgGluA 74
   : ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
100 CAAGGACAAAGTGCCCAATGCTCCAAAACGCGTCACCAAAATCGAGCGG 149
   |||:||||:||||:||||:||||:||||:||||:||||:
74 laAlaArg....LeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAsp 89
   |||:||||:||||:||||:||||:||||:||||:||||:
150 CGGTCAAGGAGCATAAACCCCGATTGATCTACTCTGCCGGAGTCTTT... 196
   |||:||||:||||:||||:||||:||||:||||:||||:
90 PheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuG1 106
   |||:||||:||||:||||:||||:||||:||||:||||:
197 .....AATGCTCCATACGG 210

106 yGlyLysLeuLeuGluGluTyrThrGln..... 115
   ||| :|||:||||:||||:||||:||||:||||:||||:
211 CACCAAGTACTTTCCGGAGTACTCCGAGACCATTCGCGATGGGTATACCT 260
   ||| :|||:||||:||||:||||:||||:||||:||||:
116 .....LeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126
   ||| :|||:||||:||||:||||:||||:||||:||||:
261 CGCAACAGCTCTCCAATTTGGCCAGGAAGCACCAGGTGTATACATCGTGGC 310
   ||| :|||:||||:||||:||||:||||:||||:||||:
127 GlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTy 143
   ||| :|||:||||:||||:||||:||||:||||:||||:
311 GGACAACTCCTGAACCTGGCGGAAGAC.....GATGCCATCTA 348
   ||| :|||:||||:||||:||||:||||:||||:||||:
143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaThrTyrA 160
   ||| :|||:||||:||||:||||:||||:||||:||||:
349 CAACACCTGCACGGTTTGGTTCACCCCTGGTGACCTTGTGGCCCAAGCATC 398
   ||| :|||:||||:||||:||||:||||:||||:||||:
160 rGlyThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCys 176
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 GCAAGATGCATCTCTTTGACATAGATGTCGAAGGT...GGCATTGCGTTC 445
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193
   :|||:||||:||||:||||:||||:||||:||||:||||:
446 AAGAGTCGGAAACGCTGTCGCGAGGCAATGATTTCCACCATCATCAACGT 495
   ||| :|||:||||:||||:||||:||||:||||:||||:
193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210
   : ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
496 AGATGGCCACAAAGATTGGCATCGCATCTGCTACGATATTCGATTCGAGG 545
   |||:||||:||||:||||:||||:||||:||||:||||:
210 luteSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrPro 226
   |||:||||:||||:||||:||||:||||:||||:||||:
546 AGATGGCGAGGCTCTATCGCAACGCGAGGTGCGAGATGATCATCTATCCG 595
   |||:||||:||||:||||:||||:||||:||||:||||:
227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243
   : ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
596 GCTGCATTCACATGACCATGCTGCTCCTCCTGAGGAGTATTCGAGCG 645
   |||:||||:||||:||||:||||:||||:||||:||||:
243 gAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnCysG 260
   : ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
646 ATCCCGTGGCAATGACAAACCACTCTTTGTGGTCACAAACATCACCAGCC 695
   |||:||||:||||:||||:||||:||||:||||:||||:
260 lyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAsp 276
   : ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
696 GTGACACAAACGCGCGAGTATGTAGCCTATGCGCATTCATTCGTTGGTGA 745
   ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
277 ProTrpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLe 293
   ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
746 CCATGGCCCAAGGTGCAGCAGAGTGCCAGTGAAGGCGAGGAATCGTGGT 795
   |||:||||:||||:||||:||||:||||:||||:||||:
293 uAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuProV 310
   : ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
796 GGCCGATATAGATTTCTCCGAGGTGGAGCAGGTGCGTTCAGCAGATTCCC 845
   ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
310 aPheGlnHisArgArgProAspLeuTyr 319
   ||||| ||| : ||||| ||| : ||||| ||| : ||||| |||
846 TCTTTGGCAAGACGCTAGATCTGTAC 874

```

OM of: US-09-357-675C-21 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Apr 29, 2002 10:02 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09357675/runat\_29042002\_091757\_3210/app\_query.fasta\_1.1208  
-DB=Issued\_Patents\_NA -OPMT=fastap -SUFFIX=p2n.rni  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -QGAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blossum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675\_CGNI\_1\_118 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-357-675C-21  
Query length: 327  
Database: Issued\_Patents\_NA:\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 85.140000

score_list:	Sequence	Strd Orig	Zscore	EScore	Len	Documentation
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-63 +			206.00	364.19	1.3e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-63 +			206.00	364.19	1.3e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-25 +			205.00	362.25	1.7e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-25 +			205.00	362.25	1.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-5 +			204.00	360.31	2.2e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-294-871A-59 +			204.00	360.31	2.2e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-5 +			204.00	360.31	2.2e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-59 +			204.00	360.31	2.2e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-61 +			203.00	358.37	2.8e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-61 +			203.00	358.37	2.8e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-9 +			202.00	356.43	3.6e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-11 +			202.00	356.43	3.6e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-21 +			202.00	356.43	3.6e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-9 +			202.00	356.43	3.6e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-9 +			202.00	356.43	3.6e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-21 +			202.00	356.43	3.6e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-917-111-1 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-479-638-1 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-1 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-3 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-7 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-13 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-15 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-17 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-19 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-23 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-27 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-29 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-31 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-47 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-49 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-51 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-55 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-65 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-1 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-3 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-7 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-13 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-15 +			201.00	354.49	4.7e-12 1785
	/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-17 +			201.00	354.49	4.7e-12 1785

/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-19 + 201.00 354.49 4.7e-12 1785  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-23 + 201.00 354.49 4.7e-12 1785  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-27 + 201.00 354.49 4.7e-12 1785  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-29 + 201.00 354.49 4.7e-12 1785  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-31 + 201.00 354.49 4.7e-12 1785  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-294-871A-63  
seq\_documentation\_block:  
: Sequence 63, Application US/08294871A  
: Patent No. 5824522  
: GENERAL INFORMATION:  
: APPLICANT: Ikenaka, Yasuhiro  
: APPLICANT: Nanba, Hirokazu  
: APPLICANT: Takano, Masayuki  
: APPLICANT: Yajima, Kazuyoshi  
: APPLICANT: Yamada, Yukio  
: APPLICANT: Takahashi, Satomi  
: APPLICANT: Okubo, Kazuma  
: APPLICANT: Yamada, Kazuhiko  
: APPLICANT: Hiraishi, Yoshiro  
: TITLE OF INVENTION: Immobilized Enzyme Preparation and  
: TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
: NUMBER OF SEQUENCES: 70  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Foley & Lardner  
: STREET: 3000 K Street, N.W.  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: U.S.  
: ZIP: 20007-5109  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/294,871A  
: FILING DATE: 22-AUG-1994  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/971,758  
: FILING DATE: 12-APR-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/917,111  
: FILING DATE: 07-AUG-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/211,641  
: FILING DATE: 11-APR-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 400848/1990  
: FILING DATE: 07-DEC-1990  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/JP91/01696  
: FILING DATE: 06-DEC-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 407922/1990  
: FILING DATE: 27-DEC-1990  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 078840/1991  
: FILING DATE: 11-APR-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 140051/1991  
: FILING DATE: 12-JUN-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/JP92/00739  
: FILING DATE: 10-JUN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 212692/1992  
: FILING DATE: 10-AUG-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PCT/JP93/01101  
: FILING DATE: 05-AUG-1993

OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

seq\_name: /cqn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-63



```

111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
464 GGTCCACTCTTTGAGAAAGCGCCGGAACCTCGGATCGGCTCAATCTGGG 513
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141
    | :::: :::: :::: :::: :::: :::: :::: :::: ::::
514 CTACGCTGAACTCGTGCGAAGAGCGGC.....GTCAAGC 548
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
549 GTCGCTTCAACACGCTCCATTTGGTGGATAAGTCAGCAAGATCGTCGGC 598
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
    ||||| :::: :::: :::: :::: :::: :::: :::: ::::
599 AAGTATCGTAAGATCCATTG.....CCGGTCCACAAG.. 631
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProv 191
    ||| :::: :::: :::: :::: :::: :::: :::: ::::
632 .....GAGTACGAGGCGCTACCGCGCTGCCAGCATCTTGAAGAAGCGTT 674
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
191 aISerThrPro.....AlaGly 196
    ||| :::: :::: :::: :::: :::: :::: :::: ::::
675 ATTTCAGCGCGGCATCTCGGCTTCCCGTCTATGACGTCGACGCGCG 724
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGlyLeuSerLe 213
    ||||| :::: :::: :::: :::: :::: :::: :::: ::::
725 AAAATGGGATGTTTCATCTGCAACGATCGCGCTGGCCTGAAGCCTGGCG 774
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheC 230
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
775 GGTGATGGGCTCTCAGGGCGCGGAGATCATCTGC.....G 809
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
230 lySerIleThrGlyProAlaHisTyrGluValLeu..... 241
    ||| :::: :::: :::: :::: :::: :::: :::: ::::
810 GCGGTACAACACGCGCGCCCAACATCCCATGTTCCCCAGCAGCACCC 859
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
242 .....LeuArgAlaArgAlaIleGluTh 249
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
860 CTGACGCTCTTCCACCACATCTCTATCGATCGAGCGCGGTCTTATCAGAA 909
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
910 CGGGCCCTGTGTCGGCGCGCGGGAAGTGGGCATG...GAGGAGAACT 956
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
266 rgAlaSerTyrGlyHisSerMetValAlaAspProTrpGlyThrValVal 282
    ||||| :::: :::: :::: :::: :::: :::: :::: ::::
957 GCATGCTGTCGGCCACTCTCTCATCTGTGGCGCCGCGGGGAATCGTC 1006
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
    ||| :::: :::: :::: :::: :::: :::: :::: ::::
1007 GCTCTCACTACGACGCTGGAAGACGAGGTGATCACCGCGCGCTCGATCT 1056
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
    |::: ||||| :::: :::: :::: :::: :::: :::: |||||
1057 CGATCGCTGCGGGAACTCGTGGAACACATCTTCAACTTCAAGCAGCATC 1106
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
314 rgArgProAspLeuTyrGly 320
    ||||| :::: :::: :::: :::: :::: :::: :::: ::::
1107 GTACGCCCCAGCATGTGT 1126
    :::: :::: :::: :::: :::: :::: :::: :::: ::::
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-25
seq_documentation_block:
; Sequence 25, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanka, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Tomomi
; APPLICANT: Okubo, Kazuma

```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-25
seq_documentation_block:
: Sequence 25, Application US/08294871A
: Patent No. 5824522
: GENERAL INFORMATION:
: APPLICANT: Ikenaka, Yasuhiro
: APPLICANT: Namba, Hirokazu
: APPLICANT: Takano, Masayuki
: APPLICANT: Yajima, Kazuyoshi
: APPLICANT: Yamada, Yukio
: APPLICANT: Takahashi, Satomi
: APPLICANT: Okubo, Kazuma
```

APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
ORIGINAL SOURCE:  
STRAIN: JM109 pAD445

FEATURE:  
NAME/KEY: CDS  
LOCATION: 233..1144  
US-08-294-871A-25  
alignment\_scores:  
Quality: 205.00 Length: 335  
Ratio: 1.273 Gaps: 15  
Percent Similarity: 48.060 Percent Identity: 25.970  
alignment\_block:  
US-09-357-675C-21 x US-08-294-871A-25 ..  
Align seg 1/1 to: US-08-294-871A-25 from: 1 to: 1785  
50 AlaValCysGlnValThr.....SerThrProAs 59  
||||| : : : : :  
164 GCGATGTGTTTCACACAGTTTCCCGCGCGCTGGCGCGACATCACCTAG 213  
59 pLysGlnAsnPheLysThr..... 66  
||||| : : : : :  
214 GAAGGAGCAGAGGTTTCATCACACGTCAGATGATCTTGCAGTGGGACAAC 263  
67 .....Cys 67  
264 AAGTCCGATCCGCGCGCGGAGACACGACAGGTCGTCGTCGTCCTT 313  
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84  
: : : : : : : : : : :  
314 CTCGACATGCTGCGAAAGCGCGGAGCGCGGCGGAATTCATTGCTCTT 363  
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95  
: : : : : : : : : : :  
364 CCCGAACTCGCGCTTACGACCTTCTCCCGCGTGGCATTTTCCACGACG 413  
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110  
||||| : : : : :  
414 AGCGCGAGCTCGATAGCTTATGAGACCGAATGCCCGCGCGGTGGTC 463  
111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127  
: : : : : : : : : : :  
464 CGTCCACATCTTTGAGAAGCGCGGAACTCGGATCGGCTCAATCTGGG 513  
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141  
: : : : : : : : : : :  
514 CTACGCTGAACCTCGTCGTCGAAGCGGC.....GTCAAGC 548  
141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAla 157  
: : : : : : : : : : :  
549 GTCGCTTCAACACAGTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGC 598  
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174  
||||| : : : : :  
599 AAGTATCGTAAGATCATTTG.....CCGGGTCAACAAG... 631  
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191  
: : : : : : : : : : :  
632 .....GAGTACGAGGCGCTACCGCGCGTTCACGATCTTGAAGACGTT 674  
191 aIleSerThrPro.....AlaGly 196  
: : : : : : : : : : :  
675 ATTTCGAGCGCGCGATCTCGGCTTCCCGGTCTATGACGTCGACGCGCGG 724  
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213  
: : : : : : : : : : :  
725 AAATGGGGATGTTTCATCTGCAACGATCGCGCTGGCCTGAAGCCTGGG 774  
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThr.....TyrProSerA 228  
: : : : : : : : : : :  
775 GGTGATGGGCGCTCAGGGCGCGGAGATCATCTGCGGCGGCTACACACGC 824  
228 laPheGlySerIleThrGlyProAlaHisTrpGluValLeu..... 241

```

; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRAND: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD445
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
; US-08-876-398A-25

alignment_scores:
    Quality: 205.00      Length: 335
    Ratio: 1.273        Gaps: 15
    Percent Similarity: 48.060      Percent Identity: 25.970

alignment_block:
    US-09-357-675C-21 x US-08-876-398A-25 ..

Align seg 1/1 to: US-08-876-398A-25 from: 1 to: 1785

50 AlaValCysGlnValThr.....SerThrProAs 59
   |||:||| ||| ||| ||| ||| ||| ||| |||
164 GCGATGTTTTCACACAGCTTTCCCGCGCGTGGCGCGACATCACCTAG 213
   |||:||| ||| ||| ||| ||| ||| ||| |||
59 pLysGlnGlnAsnPhelysthr..... 66
   |||:||| ||| ||| ||| ||| ||| ||| |||
214 GAAGGAGCAGAGTTTCATGACACGTCAGATGATCTTGCAGTGGGACAAC 263
   |||:||| ||| ||| ||| ||| ||| ||| |||
67 .....Cys 67

264 AGTCTCGATCGCGCGCGGAGACGCGAACAGTCTGCTGCTGCTT 313
   |||:||| ||| ||| ||| ||| ||| ||| |||
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
   |||:||| ||| ||| ||| ||| ||| ||| |||
314 CTCGACATGCTGACGAAAGCGCGCGCGCGGCGGATTTTCATTGCTT 363
   |||:||| ||| ||| ||| ||| ||| ||| |||
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
   |||:||| ||| ||| ||| ||| ||| ||| |||
364 CCCGAACTCGCGCTTACGACCTTTCTTCCGCGCTGGCATTTCCACGACG 413
   |||:||| ||| ||| ||| ||| ||| ||| |||
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
   |||:||| ||| ||| ||| ||| ||| ||| |||
414 AGCGGAGCTCGATAGTCTTATGAGACCGGAAATGCCCGCGCGTGGTC 463
   |||:||| ||| ||| ||| ||| ||| ||| |||
111 GluGluThrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
   |||:||| ||| ||| ||| ||| ||| ||| |||
464 CGTCCACTCTTTGAGAAGCGCGCGGAACTCGGATCGCTCAATCGGG 513
   |||:||| ||| ||| ||| ||| ||| ||| |||
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141
   |||:||| ||| ||| ||| ||| ||| ||| |||
514 CTACGCTGAACCTCGTCGTCGAAGGCGGC.....GTCAAGC 548
   |||:||| ||| ||| ||| ||| ||| ||| |||
141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaAla 157
   |||:||| ||| ||| ||| ||| ||| ||| |||
549 GTCGCTTCAACACGCTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGGC 598
   |||:||| ||| ||| ||| ||| ||| ||| |||
158 ThrTrpArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
   |||:||| ||| ||| ||| ||| ||| ||| |||
599 AAGTATCGTAAATCCATTG.....CCGGTCCACAAG.. 631
   |||:||| ||| ||| ||| ||| ||| ||| |||
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
   |||:||| ||| ||| ||| ||| ||| ||| |||
632 .....GAGTACGAGGCGCTACCGCGCGCTTCCAGCATCTTTGAAAGCGTT 674
   |||:||| ||| ||| ||| ||| ||| ||| |||
```

```

; Sequence 25, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satoshi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
```

```
191 alserThrPro.....Alaaly 196
    |||
675 ATTCGAGCGGGGATCTCGGCTTCCGGTCTATGAGTCGACGCCGCG 724
    |||
197 LysileGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
    |||
725 AAAATGGGATGTTTCATCTCAGACGATCGCGCTGGCTGAAGCTGGCG 774

213 uAlaLeuAlaGlnAlaGlyAlaGluLeuThr.....TyrProSerA 228
    :
775 GGTGATGGGCTCAGGGCGCGGAGATCATCTGCGGCGGCTACACACGC 824

228 laPheGlySerIleThrGlyProAlaHisTrpGluValLeu..... 241
    :
825 CGACCCACATCCACCGTTCGCCAGCACGACACCTGAGCTCTTCCAC 874

242 .....LeuArgAlaArgAlaIleGluThrGlnCysTyrValVa 254
    :
875 CATCTCTATCGATCGAGCGCGGCTCTATCAGAACGGGCGCTGGTCCG 924

254 lAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerTyrGlyH 271
    :
925 GCGCGGGGGAAGTGGCGATG...GAGGAGAACTGCATGCTGCTCGGCC 971

271 isSerMetValValAspProTrpGlyThrValValAlaArgCysSerGlu 287
    :
972 ACTCTGTCATCGTGGCGCGGACCGCGGGAATGCTGCTCTCACTACGAC 1021

288 ...GlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGl 303
    :
1022 CTGAAGACGAGGTGATCAGCGCGCGTGCATCTCGATCGCTGCCGCGA 1071

303 nLeuArgHisLeuProValPhe.....GlnHisArgArgProAspLeuT 319
    :
1072 ACTCGGTGACACATCTTCAACTTCAAGACGACATCGTCAGCCCGACCT 1121

319 yRGly 320
    |||
1122 ATGGT 1126

seq_name: /cqn2_5/ptodata/2/ina/5A_COMB.seq:us-08-294-871A-5

seq_documentation_block:
; Sequence 5, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
```

```

; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; STRAIN: JMI09 pad404 (FERM BP-3913)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 233..1144
; US-08-294-871A-5

alignment_scores:
    Quality: 204.00      Length: 340
    Ratio: 1.267        Gaps: 15
    Percent Similarity: 47.353      Percent Identity: 25.294

alignment_block:
US-09-357-675C-21 x US-08-294-871A-5 ..
Align seg 1/1 to: US-08-294-871A-5 from: 1 to: 1785
50 AlaValCysGlnValThr.....SerThrProAs 59
||||:||||| :|||
164 GCGATGTGTTTCAACAAGTTTTCGCCGCGCTGGCGCGGACATCACCTAG 213
```

```
59 pLysGlnAsnPhelyThr ..... 66
|||:||||:||||:
214 GAAGGACGAGGTTCTACACGCTCAGATGATCTGCAGTGGGACAAC 263
67 .....Cys 67
264 AAGTCCGATCGCGCGCGGACGACGCGAAGATCGTCTGTCGTCTT 313
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
:||||:||||:||||:
314 CTCACATGCTGACGAAGCGCGGAGCGCGCGGCGGAAATTCATTGCTT 363
84 uProGlu.....AlaPheAspPheLeAlaArg.....AspP 95
:||||:||||:||||:
364 CCGCGAAGCTCGCGCTTACGACCTTCTCCCGCGCTGCATTTCCACGACG 413
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
|||||:||||:||||:
414 AGCGCGAGCTCGATAGCTTCTATGAGACCGAATGCGCGCGCGGTGTC 463
111 GluGluThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG1 127
:||||:||||:||||:
464 CGTCCACTCTTGAGAGCGCGGAACTCGGATCGGCTTCAATCTGGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141
:||||:||||:||||:
514 CTACGCTGAAGCTCGCTCGAAGCGGC.....GTCAAGC 548
141 ySileTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValVala 157
|||||:||||:||||:
549 GTCGCTTCAACAGCTCCATTTTGTGTGATGATGATGAGCAAGATCGTCGGC 598
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
|||||:||||:||||:
599 AAGTATCGTAAGATCCATTG.....CCGGGTCAAG.. 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
|||:||||:||||:
632 .....GAGTACGAGCGCTACGCGCGCTCCAGCATCTTGAAAGCGTT 674
191 alSerThrPro.....AlaGly 196
|||:||||:||||:
675 ATTTTCGAGCGCGGATCTCGGCTTCCGCTCTATGATGATGCGCGCG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
|||:||||:||||:
725 AAATGGGATGTCATCTGCAAGCATCGCGCTGGCTGAAGCCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
:||||:||||:||||:
775 GGTGATGGCTCAGGCGCGCGAGATCATCTGC.....G 809
230 lySerIleThrGlyProAlaHisTrpGluValLeu..... 241
|||:||||:||||:
810 GCGCTACAACACCGCGACCCACCAATCCCTGTGTCCCGACGACGACC 859
242 .....LeuArgAlaArgAlaIleGluTh 249
860 CTGAGTCTTCCACCATCTCTATCATGATGAGCGCGGTCTTATCAGAA 909
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGlyLysA 266
:||||:||||:||||:
910 CGGGCGCTGGTCCGCGCGCGGCGGAGTGGGATG...GAGGAGACT 956
266 rgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVal 282
|||||:||||:||||:
957 GCATGCTGCTGCGCACTTCGATCGTGGCGCGCGGGAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
|||:||||:||||:
1007 GCTCTACTAGCGCTGGAAGACGAGGTGATCATCCGCGCGCGTCTGATCT 1056
```

```
298 uAsnTyrLeuArgGlnLeuArgHisLeuProValPhe...GlnHisA 314
|||:||||:||||:
1057 CGATGCTGCGGGAAGTGGTGAACACATCTTCAACTTCAAGCAGCATC 1106
314 rgArgProAspLeuTyrGly 320
|||:||||:||||:
1107 GTCAGCCCGACGACTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-59

seq_documentation_block:
; Sequence 59, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; PROCESS OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
```

Mon Dec 16 10:44:58 2002

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 PAD404 (FERM BP-3913)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(233..1141)
; US-08-876-398A-5

alignment_scores:
  Quality: 204.00      Length: 340
  Ratio: 1.267        Gaps: 15
  Percent Similarity: 47.353  Percent Identity: 25.294

alignment_block:
US-09-357-675c-21 x US-08-876-398A-5  ..
Align seg 1/1 to: US-08-876-398A-5 from: 1 to: 1785

50 AlaValCysGlnValThr.....SerThrProAs 59
||||:||||:||||:||||:||||:||||:||||:
164 GCGATGTTTCAACAGCTTTTCCCGCGCGTGGCGGACATCACCTAG 213
59 pLysGlnGlnAsnPhelYsthr..... 66
||||:||||:||||:||||:||||:||||:
214 GAAGGACGAGGGTTTCATGACACGTCAGATCATCTTGCAGTGGGACAAC 263
67 .....Cys 67
264 AAGGTCGGATCGCGCGCGGAGACGCGAACAGGTCGTGCTGCTT 313
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
||||:||||:||||:||||:||||:||||:
314 CTCACATGCTGACGAAGCGCGCGCGGCGCGGAATTTCATTGCTT 363
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95
||||:||||:||||:||||:||||:||||:
364 CCCGAACTCCGGCTTAGACCTTCTCCCGCGGTGGCATTTCCACGACG 413
95 roAlaGluThrLeuHisSerGlu...ProLeuGlyGlyLeuLeu 110
||||:||||:||||:||||:||||:||||:

```

```

414 AGCCCGAGCTCGATAGCTTCTATGAGACCAGAAATGCCCGCCCGGTGGTC 463
111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG1 127
::: ::: ||| ||| |||:||||:||||:
464 CGTCCACTCTTTGAGAAGCGCGGAACTCGGATCGGCTTCAATCTGGG 513
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141
| ::: ||| |||
514 CTACGCTGAACCTCGTCGTCGAAGCGGC.....GTCAAGC 548
141 ySileTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValVala1 157
:::||||:||||:||||:||||:||||:
549 GTCGCTTCAACACGCTCCATTTGGTGGATAAGTCAGGCAAGATCGTCGC 598
158 ThrTyrArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPr 174
||||:||||:||||:||||:||||:
599 AAGTATCGTAGATCCATTG.....CCGGTCAACAAG.. 631
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191
||| :::: ||| |||:||||:
632 .....GAGTACGAGGCTTACCGCGGCTCCAGCATCTTGAAAGCGTT 674
191 alSerThrPro.....AlaGly 196
||| |||:||||:
675 ATTTGAGCGCGCGGATCTCGGCTTCCGCTCTATGACGTCGACGCGCG 724
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
||||:||||:||||:||||:||||:||||:
725 AAATGGGATGTTCACTGCAAGATCGCGCTGGCTGAAGCCTGGCG 774
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
:::||||: ||| ||| |||:||||:
775 GTGATGGGCTCAGGGGCGCGAGATCATCTGC.....G 809
230 lySerIleThrGlyProAlaHisTrpGluValLeu..... 241
||||: ||| |||:||||:
810 GCGGCTACAACGCGCGACCCCAATCCCTTGTCCCGACGACGACAC 859
242 .....LeuArgAlaAlaGluIleGluTh 249
860 CTGACGCTCTCCACATCTCCTATCGATCGAGCGCGGCTTTATCAGAA 909
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGluLysA 266
: ::::: ||| |||:||||: |||
910 CGGGCCTGTTCCGCGCGCGCGGCAAGTGGGATG...GAGGAACT 956
266 rGAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
||||:||||: ||| ||| ||| |||
957 GCATGCTGCTCGCCACTCTCTGATCGTGGCGCGACCGGGAATGCT 1006
283 AlaArgCysSerGlu...GlyProGlyLeuAlaAlaArgIleAspLe 298
||| |||:||||: ||| ||| |||:||||:
1007 GCTCTACTACGAGGTGGAGAGAGGATGATCACCGCGCGCTGATCT 1056
298 uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
::: |||:||||:||||: ||| ||| |||
1057 CGATCGTCCGGGAACCTCGTGAACACATCTTCAACTTCAAGCAGCATC 1106
314 rArgProAspLeuTyrGly 320
||||:||||: ||| |||
1107 GTCAGCCCGACCATATGTT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-876-398A-59
seq_documentation_block:
; Sequence 59, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAMBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio

```





```
|||||
1107 GTCAGCCCCGACCTATGGT 1126

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-294-871A-61

seq_documentation_block:
; Sequence 61, Application US/08294871A
; Patent No. 5824522
;
GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshiro
;
TITLE OF INVENTION: Immobilized Enzyme Preparation and
;
NUMBER OF SEQUENCES: 70
;
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
;
ZIP: 20007-5109
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
SOFTWARE: Patent In Release #1.0, Version #1.25
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
;
CLASSIFICATION: 435
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
;
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
;
INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; STRAIN: JMI09 PAD468
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 233..1144
US-08-294-871A-61

alignment_scores:
Quality: 203.00 Length: 340
Ratio: 1.261 Gaps: 15
Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
US-09-357-675C-21 x US-08-294-871A-61 ..

Align seg 1/1 to: US-08-294-871A-61 from: 1 to: 1785

50 AlaValCysGlnValThr.....SerThrProAs 59
|||||
164 GCGATGTTTCACACGTTTCCCGCCCTGGCGGACATCACCTAG 213
|||||
59 pLysGlnGlnAsnPhelLysThr..... 66
|||||
214 GAAGGACGAGAGTTTCATGACAGTCAGATGATACATTGCGAGTGGGACAAC 263
|||||
67 .....Cys 67

264 AAGGTCGGATCGCGCGCGGAGACACGCGAGTCGTCGTTGCTCTT 313
|||||
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
|||||
314 CTGACATGCTGAGAAAGCGCGCGCGGCGGAATTTCATTGCTCTT 363
|||||
84 uProGlu.....AlaPheAspPheLeAlaArg.....Asp 95
|||||
364 CCCGGAACCTCGCGCTTACGACCTTCTCCCGCGCTGCGCATTTACCCAGC 413
|||||
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
|||||
414 AGCGCAGCTCGATAGCTTCTATGACACCGAAATGCCCGCGCGGTGTC 463
|||||
111 GluGluTyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG 127
|||||
464 CGTCCACTCTTTGAGAAGCGCGGAACTCGGATCGGCTTCAATCTGGG 513
|||||
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141
|||||
514 CTACGGTGAACCTGCTCGTGAAGCGGC.....GTCAAGC 548
|||||
141 ysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157
|||||
549 GTCGCTTCAACAGCTCCATTTTGTGGATAAGTCAGGCAAGATCGTCGC 598
|||||
158 ThrTyArgLysThrHisLeuCysAspValIleProGlyGlnGlyPr 174
|||||
599 AAGTATCGTAAGATCCATTG.....CCGGTCAACAAG.. 631
|||||
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerPro 191
```

```

632      |||| : : : : : |||| ||||| : : :
      .....GAGTACGAGCCCTACCGCGGTTCACAGCATCTTGAAGAGCGTT 674

191  alSerThrPro.....|||
      |||
675  ATTTCAGCGCGCGATCTCGGCTTCGCCGTCTATGAGTCGACGCCGCG 724

197  LysilecylLeuAlaValcysTyrAspMetArgPheProCluLeuSerLe 213
      ||||| : : : : : |||| : : : : : ||||| : : : : :
725  AAAATGGGATGTCATCTGCAACGATCGCGCTGGCTTGAAGCCCTGGCG 774

213  uAlaLeuAlaGlnAlaGlyAlaGluLeuLeuThrTyrProSerAlaPheG 230
      : : : : : ||||| : : : : :
775  GGTGATGGGCTCAGGCGCCGAGATCATCTGC.....G 809

230  lysSerIleThrGlyProAlaHisTrpCluValLeu..... 241
      ||| : : : ||| : : : : : : : : :
810  GCGGCTACAAACAGCCGACCAATCCCGCTGTCCCCAGCAGACCCAC 859

242  .....LeuArgAlaArgAlaIleGluTh 249
      : : : : : ||| : : : : :

860  CTGAGGTCCTTCACCATCTCTTCATCGATCAGSCCGGGTCTTATCAGAA 909

249  rGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGluLysA 266
      : : : : : ||| : : : : : |||
910  CGGGGCTGTGTCGCGCGCGCGGGAAGTGGCGATG...GAGAGAACT 956

266  rGAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
      ||||| : : : : : ||||| : : : : :
957  GCATGCTGCTCGGCCACTCTGTGATCGTGGCGCGGAGGAAATCGTC 1006

283  AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
      ||| : : : : : ||| : : : : :
1007  GCTCTCACTACGAGCTGGAAGACGAGGTGATCACGCGCGCGCTCATCT 1056

298  uAsnTyrLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisA 314
      : : : : : ||||| : : : : : ||||| : : : : :
10057  CGATCGTGCCGGAACTGCGTGAACACATCTTCACTTCAACGACGATC 1106

314  rGArgProAspLeuTyrGly 320
      ||| : : : : : |||||
107  GTCAGCCCCAGCACTATGTT 1126

```

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-876-398A-61

```

seq_documentation_block:
; Sequence 61, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U. S. A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997

```

```

> CLASSIFICATION: 435
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: US 08/211,641
> FILING DATE: 11-APR-1994
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: WO PCT/JP93/01101
> FILING DATE: . 05-AUG-1993
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: JP 340078/1992
> FILING DATE: 21-DEC-1992
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: JP 212692/1992
> FILING DATE: 10-AUG-1992
> ATTORNEY/AGENT INFORMATION:
> NAME: Wegner, Harold C.
> REGISTRATION NUMBER: 25,258
> REFERENCE/DOCKET NUMBER: 74129/130
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (202) 672-5300
> TELEFAX: (202) 672-5399
> INFORMATION FOR SEQ ID NO: 61:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1785 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: double
> TOPOLOGY: linear
> MOLECULE TYPE: DNA (genomic)
> ORIGINAL SOURCE:
> ORGANISM: Escherichia coli
> STRAIN: JMI09 PAD468
> FEATURE:
> NAME/KEY: CDS
> LOCATION: join(233..1141)
> US-08-876-398A-61

alignment_scores:
    Quality: 203.00      Length: 340
    Ratio: 1.261         Gaps: 15
Percent Similarity: 47.353      Percent Identity: 25.294

alignment_block:
US-09-357-675C-21 x US-08-876-398A-61      ..

Align seg 1/1 to: US-08-876-398A-61 from: 1 to: 1785

50 AlaValCysGlnValThr.....SerThrProAsn 59
   |||||:|||||:|||||:|||||:|||||:|||||
164 GCGATGTGTTTCAACAGCTTTTCCGCGCGCTGGGCGGACATCACCTAG 213
   59 playsGlnGlnAspPheLysThr..... 66
   |||||:|||||:|||||:|||||:|||||:|||||
214 GAAGGAGCAGAGGTTCTATGACAGCTCAGATGATACTTCCAGTGGGACAC 263
   67 .....Cys 67

264 AAGGTCGATCGCGCGCGGAGACACCGAACAGGTCGTCGTCGTCCT 313

68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
314 CTCGACATGCTGACGAAGCCGCGGAGCGCGGGCGGAATTTTCATTGTCT 363
   84 uProGlu.....AlaPheAspPheLeAlaArg.....AspP 95
   :|||||:|||||:|||||:|||||:|||||:|||||
364 CCCCAGAACTCGCGCTTACGACCTTCTTTCCGCGCTGGCATTTTCACCGAG 413

95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
   |||||:|||||:|||||:|||||:|||||:|||||:|||||
414 AGGCCGAGCTCGATAGCTTCTATGAGACCGGAATTCGCCGCGGTGGTC 463
   |||||:|||||:|||||:|||||:|||||:|||||

111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG 127
   :|||:|||||:|||||:|||||:|||||:|||||:|||||
464 CGTCCACTCTTTCAGAGGCGCGGAGAACTCGGGATCGGCTTCAATCTGGG 513
   :|||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-294-871A-9

**C**



APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
PROCESS FOR PRODUCING D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294.871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1785 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear  
ORIGINAL SOURCE:  
STRAIN: JM109 PAD439  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 233..1144  
US-08-294-871A-21  
alignment\_scores:  
Quality: 202.00 Length: 340  
Ratio: 1.255 Gaps: 15  
Percent Similarity: 47.353 Percent Identity: 25.294  
alignment\_block:  
US-09-357-675C-21 x US-08-294-871A-21  
Align seg 1/1 to: US-08-294-871A-21 from: 1 to: 1785  
50 AlaValCysGlnValThr.....SerThrProAs 59  
|||||.....  
164 GCGATGTGTTTCAACAGTTTCCCGCGCGTGGCGCGACATCACCTAG 213  
59 pLysGlnGlnAsnPheLysThr..... 66  
|||||.....  
214 GAAGAGCAGAGGTTTCATGCACGTCAGATGATCTGTCAGTGGGACAAC 263  
67 .....Cys 67  
264 AAGGTCGATCGCGCGCGAGACACGCGAGGTCGTCGTCGTCCTT 313  
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84  
.....  
314 CTCGACATCTGACGAAACCGCGCGCGGCGGCGAATTTTCATTGTCCTT 363  
84 uProGlu.....AlaPheAspPheIleAlaArg.....Asp 95  
|||||.....  
364 CCCGAACTCGCGCTTACGACCTTCTCCCGCGCTGGCATTTCCACGCGAG 413  
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110  
|||||.....  
414 AGCGCGAGCTCGATAGCTTCTATGAGACCGAAATGCCGCGCGGTGTC 463  
111 GluGluTyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127  
.....  
464 CGTCCACTCTTTGGAAGCGCGGAACTCGGGATCGGCTTCAATCTGGG 513  
127 yGlyPheHis.....GluArgGlyGlnAspTrpGluGlnThrGlnL 141  
|||||.....GTCAAGC 548  
514 CTACGCTGAACCTCGTCGTCGAAGCGGC.....  
141 ysIleTyrrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAla 157  
.....  
549 GTCGCTTCAACACAGCTCCATTTTGGTGGATAAGTCAGGCAAGATCGTCGC 598  
158 ThrTyrrArgLysThrHisLeuCysaspValGluIleProGlyGlnGlyPr 174  
|||||.....  
599 AAGTATCGTAAGATCCATTG.....CCGGGTCAACAAG... 631  
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProV 191  
|||||.....  
632 .....GAGTACGAGGCCCTACCGCGCTTCCAGCATCTTGAAGCGGTT 674  
191 alSerThrPro.....AlaGly 196  
|||||.....  
675 ATTTCGAGCGCGCGATCTCGGCTTCCCGCTCTATGACGTCGACGCGCGG 724  
197 LysIleGlyLeuAlaValCysTyrrAspMetArgPheProGluLeuSerLe 213  
|||||.....  
725 AAAATGGGGATGTTTCATCTGCAACGATCCCGCTGCGCTCGAAGCCCTGGCG 774  
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrrProSerAlaPheG 230

```

775 GGTGATGGCGCTCAGGGCCGCAGATCATCTGC.....|
230 lySerIleThrlyProAla.....|
810 CGGGCTACAACAGCGCGCACCAATCCCCCTGTTCCGCCAGCAGACCAC 859
237 .....HisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
860 CTGAGCTCTTTCACCATCTCTCATCGATGCAGGCGGGTCTATTACAGAA 909
249 rGlnCysYrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysA 266
910 CGGGCCCTTGTCGGCGCGCGGAAGCTGGCATG...GAGGAGAAGCT 956
266 rgAlaSerTyrglyHisSerMetValValAspProTrpGlyThrValVal 282
957 GCATCGTCTCGGCCCATCTCTCATCGTGGCGCCGACCGGGGAAATCGTC 1006
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleaspLe 298
1007 GCTCTCACTACAGCGCTGGAAGACGAGGTGATCACCGCGCGCTCGATCT 1056
298 wasnTyrlLeuArgGlnLeuArgArgHisLeuProValPhe...GlnHisa 314
1057 GCATCGTCCGGGGAACTCGTGAACACATCTTCAACTTCAAGCAGCATC 1106
314 rgArgProAspLeuTyrgly 320
1107 GTACCCCCAGCATATGTT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-9
seq_documentation_block:
; Sequence 9' Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992

```

```
174 oMetCysGluSerAsnSerThrMetProGlyProSerLeuGluSerProv 191
    ||| : : : : |||
    ||| : : : : |||
632 .....GAGTACGAGCCTACCGCGGTTCACGATCTTGAAAGCGTT 674
    ||| : : : : |||
191 alSerThrPro.....AlaGly 196
    |||
675 ATTTGAGCGCGCGATCTCGGCTTCCCGTCTATGACGTCGACGCGCG 724
    |||
197 LysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLe 213
    ||| : : : : |||
725 AAAATGGGATGTCATCTGCAACATGCGCGCTGGCTGCAAGCCTGGCG 774
    ||| : : : : |||
213 uAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheG 230
    : : : : : ||| : : : : |||
775 GGTGATGGGCTCAGGGCGCGGATCATCTGC.....G 809
    ||| : : : : |||
230 lySerIleThrGlyProAla..... 236
    ||| : : : : |||
810 GCGGCTACAAACAGCGCGGACCAATCCCTGTTCCCGACGACGACCAC 859
    ||| : : : : |||
237 .....HisTrpGluValLeuLeuAlaArgAlaIleGluTh 249
    ||| : : : : |||
860 CTGAGTCTCTCCACCATCTCTATCGATCGAGCGCGGTCTTATCAGAA 909
    ||| : : : : |||
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGluLysA 266
    : : : : : ||| : : : : |||
910 CGGGGCTGTGTCGCGCGCGCGGCAAGGCGGCATG...GAGGAGAACT 956
    ||| : : : : |||
266 rGlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
    ||| : : : : |||
957 GCATGTGTCTCGGCCACTCTGTCATCTGGCGCGCACCGGGAAATCGTC 1006
    ||| : : : : |||
283 AlaArgCysSerGlu...GlyProGlyLeuCysLeuAlaArgIleAspLe 298
    ||| : : : : |||
1007 GCCTCTACTACGACGCTGGAAGAGGAGGTGATCACC CGCGCGCTCGATCT 1056
    ||| : : : : |||
298 uAsnTyrLeuArgGlnLeuArgHisLeuProValPhe...GlnHisA 314
    ||| : : : : |||
1057 CGATCGCTCGCGGAACTCGGTCAACACATCTTCAACTTCAACGACGATC 1106
    ||| : : : : |||
314 rArgProAspLeuTyrGly 320
    ||| : : : : |||
1107 GTCAGCCCGACGACTATGTT 1126

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-876-398A-11

seq_documentation_block:
; Sequence 11, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1785 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: JM109 pad428
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(233..1141)
;
US-08-876-398A-11
```

```
alignment_scores:
    Quality: 202.00      Length: 340
    Ratio: 1.255        Gaps: 15
Percent Similarity: 47.353 Percent Identity: 25.294

alignment_block:
US-09-357-675C-21 x US-08-876-398A-11 ..

Align seg 1/1 to: US-08-876-398A-11 from: 1 to: 1785

50 AlaValCysGlnValThr.....SerThrProAs 59
    ||| : : : : |||
164 GCGATGTGTTCACAAGCTTTTCCGCGCGCTGGCGCGGACATCACCTAG 213
    ||| : : : : |||
59 pLysGlnGlnAsnPheLysThr.....Cys 66
    ||| : : : : |||
214 GAAGGAGCAGAGGTTTCATGCACGTCAGATGATACTTGCAGTGGGACAAC 263
    ||| : : : : |||
67 ..... 67
264 AGGTCCGATCGCGCGGAGACACGCGACAGGTCGTCGTCGTCCT 313
    ||| : : : : |||
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84
    : : : : : ||| : : : : |||
314 CTCGACATGCTGACGAAGCCGCGAGCGCGGCGGAATTTTCATTGCTCTT 363
    ||| : : : : |||
84 uProGlu.....AlaPheAspPheIleAlaArg.....AspP 95
    : : : : : ||| : : : : |||
364 CCCCGAACTCGCGCTTACGACCTTCTCCCGCGCTGGCATTTTCCCGGAG 413
    ||| : : : : |||
95 roAlaGluThrLeuHisLeuSerGlu...ProLeuGlyGlyLysLeuLeu 110
    ||| : : : : |||
414 AGCCGCGCTCGATAGCTTCTATGAGACCGGAATGCCCGCGCGGTGGTC 463
    ||| : : : : |||
111 GluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127
```







OM of: US-09-357-675C-21 to: Pending\_Patents\_NA\_Main:\* out\_format : pfs

Date: Apr 29, 2002 12:18 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09357675/runat\_29042002\_091758\_3302/app\_query.fasta\_1.1208  
-DB=Pending\_Patents\_NA\_Main -OPMT=fastap -SUFFIX=p2n.rnps  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOPEXT=0.000 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000  
-XGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blossum62 -TRANS=human40.cdi -LIST=45 -DOALIGN=200  
-THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09357675 -CGN1\_10416 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-357-675C-21

Query length: 327

Database: Pending\_Patents\_NA\_Main: \*

Database sequences: 21979536

Database length: 2067452561

Search time (sec): 4349.110000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/pna/US096C.COMB.seq:US-09-652-125A-9045	1	1748.00	1748.00	2640.38	1.3e-137
/cgn2_6/ptodata/2/pna/US097B.COMB.seq:US-09-726-802-2166	1	1748.00	1748.00	2640.38	1.3e-137
/cgn2_6/ptodata/2/pna/US0821.COMB.seq:US-09-213-360-100	1	1743.00	1743.00	2632.52	3.5e-137
/cgn2_6/ptodata/2/pna/US0931.COMB.seq:US-09-357-675A-1	1	1743.00	1743.00	2635.19	2.5e-137
/cgn2_6/ptodata/2/pna/US0933.COMB.seq:US-09-357-675C-1	1	1743.00	1743.00	2635.19	2.5e-137
/cgn2_6/ptodata/2/pna/US0627.COMB.seq:US-06-278-258-200	1	1730.00	1730.00	2607.89	8.3e-136
/cgn2_6/ptodata/2/pna/PCRTUS.COMB.seq:PCT-US00-05988-692	1	1718.00	1718.00	2597.42	3.2e-135
/cgn2_6/ptodata/2/pna/US099A.COMB.seq:US-09-925-300-692	1	1718.00	1718.00	2597.42	3.2e-135
/cgn2_6/ptodata/2/pna/US0617.COMB.seq:US-06-172-373-14716	1	1718.00	1718.00	2598.13	1.5e-134
/cgn2_6/ptodata/2/pna/US0616.COMB.seq:US-06-164-285-3819	1	1717.00	1717.00	2585.99	1.4e-134
/cgn2_6/ptodata/2/pna/US0933.COMB.seq:US-09-359-922-12822	1	1626.50	1626.50	2448.90	5.9e-127
/cgn2_6/ptodata/2/pna/US0933.COMB.seq:US-09-359-922-12822	1	1626.50	1626.50	2448.90	5.9e-127
/cgn2_6/ptodata/2/pna/US0636.COMB.seq:US-06-360-207-3686	1	1487.00	1487.00	2246.87	1.1e-115
/cgn2_6/ptodata/2/pna/US0624.COMB.seq:US-06-242-679-78	1	1468.50	1468.50	2208.98	1.4e-113
/cgn2_6/ptodata/2/pna/US0632.COMB.seq:US-06-324-185-10674	1	1315.00	1315.00	1971.69	2.3e-100
/cgn2_6/ptodata/2/pna/PCRTUS.COMB.seq:PCT-US98-16235-92	1	1067.00	1067.00	1609.11	3.5e-80
/cgn2_6/ptodata/2/pna/US092.COMB.seq:US-09-244-112-92	1	1067.00	1067.00	1609.11	3.5e-80
/cgn2_6/ptodata/2/pna/US097C.COMB.seq:US-09-774-639-92	1	1067.00	1067.00	1609.11	3.5e-80
/cgn2_6/ptodata/2/pna/US099C.COMB.seq:US-09-969-730-76	1	1067.00	1067.00	1609.11	3.5e-80
/cgn2_6/ptodata/2/pna/US095B.COMB.seq:US-09-535-896-18207	1	994.00	994.00	1504.83	2.3e-74
/cgn2_6/ptodata/2/pna/US098B.COMB.seq:US-09-855-294B-10	1	968.50	968.50	1459.29	7.8e-72
/cgn2_6/ptodata/2/pna/US0933.COMB.seq:US-09-354-899-1617	1	873.00	873.00	1320.80	4.1e-64
/cgn2_6/ptodata/2/pna/US096C.COMB.seq:US-09-652-911-7920	1	845.50	845.50	1271.87	2.2e-61
/cgn2_6/ptodata/2/pna/US095B.COMB.seq:US-09-535-896-18219	1	801.00	801.00	1212.83	4.2e-58
/cgn2_6/ptodata/2/pna/US092.COMB.seq:US-09-235-076-3825	1	794.00	794.00	1202.82	1.5e-57
/cgn2_6/ptodata/2/pna/US092.COMB.seq:US-09-289-768-35986	1	794.00	794.00	1202.82	1.5e-57
/cgn2_6/ptodata/2/pna/US0933.COMB.seq:US-09-332-782-33825	1	794.00	794.00	1202.82	1.5e-57
/cgn2_6/ptodata/2/pna/US097B.COMB.seq:US-09-737-223-23825	1	794.00	794.00	1202.82	1.5e-57
/cgn2_6/ptodata/2/pna/US099B.COMB.seq:US-09-939-397-25986	1	794.00	794.00	1202.82	1.5e-57
/cgn2_6/ptodata/2/pna/US098B.COMB.seq:US-09-834-366-11899	1	792.00	792.00	1198.45	2.6e-57
/cgn2_6/ptodata/2/pna/US0619.COMB.seq:US-06-197-873-11899	1	792.00	792.00	1198.45	2.6e-57
/cgn2_6/ptodata/2/pna/US096C.COMB.seq:US-09-652-911-9443	1	775.50	775.50	1163.12	2.5e-55
/cgn2_6/ptodata/2/pna/US097A.COMB.seq:US-09-710-282-3968	1	775.50	775.50	1163.12	2.5e-55
/cgn2_6/ptodata/2/pna/US097A.COMB.seq:US-09-716-472-5258	1	775.50	775.50	1163.12	2.5e-55
/cgn2_6/ptodata/2/pna/US092.COMB.seq:US-09-235-076-31257	1	768.50	768.50	1164.44	2.1e-55
/cgn2_6/ptodata/2/pna/US092.COMB.seq:US-09-289-768-39162	1	768.50	768.50	1164.44	2.1e-55
/cgn2_6/ptodata/2/pna/US0933.COMB.seq:US-09-332-782-31257	1	768.50	768.50	1164.44	2.1e-55
/cgn2_6/ptodata/2/pna/US097B.COMB.seq:US-09-737-223-31257	1	768.50	768.50	1164.44	2.1e-55
/cgn2_6/ptodata/2/pna/US099B.COMB.seq:US-09-939-397-39162	1	768.50	768.50	1164.44	2.1e-55

/cgn2\_6/ptodata/2/pna/US092.COMB.seq:US-09-240-371-6315 + 760.00 1151.36 1.1e-54  
/cgn2\_6/ptodata/2/pna/US092.COMB.seq:US-09-277-19834 + 760.00 1151.36 1.1e-54  
/cgn2\_6/ptodata/2/pna/US099A.COMB.seq:US-09-909-627-19834 + 760.00 1151.36 1.1e-54  
/cgn2\_6/ptodata/2/pna/US099A.COMB.seq:US-09-919-724-6315 + 760.00 1151.36 1.1e-54  
/cgn2\_6/ptodata/2/pna/US0933.COMB.seq:US-09-306-609-11407 + 759.50 1147.16 1.9e-54

seq\_name: /cgn2\_6/ptodata/2/pna/US096C.COMB.seq:US-09-652-125A-9045

seq\_documentation\_block:

Sequence 9045, Application US/09652125A

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

APPLICANT: Distefano, Peter

TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES

TITLE OF INVENTION: THEREFOR

FILE REFERENCE: 1600.1189-001

CURRENT APPLICATION NUMBER: US/09/652,125A

CURRENT FILING DATE: 2001-08-30

PRIOR APPLICATION NUMBER: 60/151,127

PRIOR FILING DATE: 1999-08-30

NUMBER OF SEQ ID NOS: 9506

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9045

LENGTH: 1826

TYPE: DNA

ORGANISM: Homo sapiens

US-09-652-125A-9045

alignment\_scores:

Quality: 1748.00 Length: 327

Ratio: 5.346 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-357-675C-21 x US-09-652-125A-9045

Align seg 1/1 to: US-09-652-125A-9045 from: 1 to: 1826

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuCy 17  
|||||

107 ATGCTGGGCTTCATCACAGGCTCTCCACAGATCTCTGCTCCCTCTGTG 156  
|||||

17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34  
|||||

157 TCCTGGACTCGGATACCTCACTCTCAGTACTTGTGCTCAGCCAGGC 206  
|||||

34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50  
|||||

207 CCAGAGCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 256  
|||||

51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelystThrCy 67  
|||||

257 GTGTGGCAGGTATACATCGAGCCGAGCAAGCAAGCAAGCAAGCAAGCA 306  
|||||

67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
|||||

307 TGCTGAGCTGGTTCGAGAGGCTCCAGACTGGGTGCTGCTGCTGCTTCC 356  
|||||

84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
|||||

357 TGCTGAGGATTTGACTTCTATTCGAGGAGCCCTGCAGAGAGCTACAC 406  
|||||

101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117  
|||||

407 CTGTCTGAACCATCTGGGTGGGAACCTTTTGGGAAGAAATACACCCAGCTGC 456  
|||||

117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134  
|||||

457 CAGGAATGTGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 506  
|||||

134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150  
|||||

507 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCCAGTGTGCTGGAAC 556  
151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167  
|||||  
557 AGCAAGGGGCAGTAGTGCCACTTACAGGAAGACACATCTGTGTACGCT 606  
167 lGluileProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
607 AGAGATTCCAGGCGCGGCTATGTGTGAAGCAACTCTACCATCCCTG 656  
184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
657 GGCCCACTCTGAGTCACTGTGACAGCAGCAGCAGCAAGATGTGCTA 706  
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
707 GCTGTCTGTATGACATGCGGTTCCTGAACTCTCTGTGATGCTCA 756  
217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234  
757 AGCTGAGCAGAGATACTTACCTATCTCTGAGCTTTTGGATCCATTACAG 806  
234 lyProAlaHisTrpGluValLeuLeuArgAlaIleGluThrGln 250  
807 GCCCAGCCCACTGGGAGGTGTGCTGCGGCGCGTGTCTATCGAAACCCAG 856  
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267  
857 TGTATGTAGTGGCAGCAGCAGCAGTGTGACGCCACCATGAGAGAGAGC 906  
267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAla 284  
907 AAGTTATGGCCACAGCATGTGTAGACCCCTGGGGAACAGTGTGGGCC 956  
284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300  
957 GCTGCTCTGAGGGGCGAGGCTCTGCTTGGCCGAATAGACCTCAACTAT 1006  
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317  
1007 CTCGACAGTGGCGGACACCTGCTGTGTTCCAGCAGCAGCGGCTCA 1056  
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327  
1057 CACTATGGCAATCTGGTCAACCCACTGTCT 1087  
seq\_name: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq:us-09-726-802-2166  
seq\_documentation\_block:  
; Sequence 2166, Application US/09726802  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L.  
; APPLICANT: Falb, Dean A.  
; APPLICANT:  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; FILE REFERENCE: 1600.2025-001  
; CURRENT APPLICATION NUMBER: US/09/726.802  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/168,012  
; PRIOR FILING DATE: 1999-11-30  
; NUMBER OF SEQ ID NOS: 2872  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2166  
; LENGTH: 1826  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-726-802-2166  
alignment\_scores:  
Quality: 1748.00 Length: 327  
Ratio: 5.346 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-357-675c-21 x US-09-726-802-2166  
Align seg 1/1 to: US-09-726-802-2166 from: 1 to: 1826  
1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17  
|||||  
107 ATGCTGGGCTTCATCACCAGGCTCCTCAGATTCCTGCTCCCTCTCTGTG 156  
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34  
|||||  
157 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTGTGCTCAGCCAGGC 206  
34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50  
|||||  
207 CCAGAGCCATGGCTATCTCTCTCTCTGCGAATGCCCTGGTGGCT 256  
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67  
|||||  
257 GTGTGCCAGGTAACTCGACGCCAGACAGCAACAACTTTAAACATG 306  
67 salaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
|||||  
307 TGTCTAGCTGGTTCGAGAGGCTCCAGACTGGGTGCTGCTGCTGCTTCC 356  
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
|||||  
357 TCCTGAGGCACTTGTGACTTTCACGGGACCTTCAGAGAGCGCTACAC 406  
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117  
|||||  
407 CTGTCTGAACCACTGGGTGGGAACTTTTGAAGAATAACACCAGCTTGC 456  
117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134  
|||||  
457 CAGGGAATGTGGACTCTGGCTGTCTTGGGTGTTTCCATGAGCGTGGCC 506  
134 InAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150  
507 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTCCAGTGTGCTGGAAC 556  
151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167  
|||||  
557 AGCAAGGGGCAGTAGTGCCACTTACAGGAAGACACATCTGTGTACGCT 606  
167 lGluileProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
|||||  
607 AGAGATTCCAGGCGCGGCTATGTGTGAAGCAACTCTACCATCCCTG 656  
184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
657 GGCCCACTCTGAGTCACTGTGACAGCAGCAGCAGCAAGATGTGCTA 706  
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
707 GCTGTCTGTATGACATGCGGTTCCTGAACTCTCTGCGATTTGGCTCA 756  
217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234  
|||||  
757 AGCTGAGCAGAGATACTTACCTATCTCTGAGCTTTTGGATCCATTACAG 806  
234 lyProAlaHisTrpGluValLeuLeuArgAlaIleGluThrGln 250  
807 GCCCAGCCCACTGGGAGGTGTGCTGCGGCGCGTGTCTATCGAAACCCAG 856  
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267  
|||||  
857 TGTATGTAGTGGCAGCAGCAGCAGTGTGACGCCACCATGAGAGAGAGC 906  
267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAla 284  
|||||

```

907 AAGTTATGGCCACAGCATGTGTGTAGACCCCTGGGGAAACAGTGGTGGCCC 956
284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgileAspLeuAsnTyr 300
957 GCTGCTCTGAGGGCCAGGCTCTGCTTGGCCGAATAGACCTCAACTAT 1006
301 LeuArgGlnLeuArgHisLeuProValPheGlnHisArgArgProAs 317
1007 CTGCGACAGTTGGCGGCACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1056
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
1057 CCTCTATGGCAATCTGGGTCAACCCACTGTCT 1087

```

seq\_name: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq:US-60-213-360-100

```

seq_documentation_block:
; Sequence 100, Application US/60213360
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor
; TITLE OF INVENTION: Identified Thereby
; FILE REFERENCE: GX-0014 P
; CURRENT APPLICATION NUMBER: US/60/213,360
; CURRENT FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 8347
; SOFTWARE: PERL Program
; SEQ ID NO 100
; LENGTH: 4174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No: 1096083.6c
; NAME/KEY: unsure
; LOCATION: 2331, 2333-2334, 2337, 2339-2340, 3582-3618
; OTHER INFORMATION: a, t, c, g, or other
US-60-213-360-100

```

```

alignment_scores:
Quality: 1748.00      Length: 327
Ratio: 5.346          Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

```

alignment\_block:

US-09-357-675C-21 x US-60-213-360-100 ..

Align seg 1/1 to: US-60-213-360-100 from: 1 to: 4174

```

1 MetLeuGlyPheLeuThrArgProHisArgPheLeuSerLeuLeuCy 17
|||||
84 ATGTCGGGCTTCATCACAGGCTCTCTCACAGATTCTGTCCTCTCTGTG 133
|||||
17 sProGlyLeuArgileProGlnLeuSerValLeuCysAlaGlnProArgp 34
|||||
134 TCTTGGACTCCGGATACCTCAACTCTCAGTACTTGTGCTCAGCCCGAGGC 183
|||||
34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50
|||||
184 CCAGAGCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 233
|||||
51 ValCysGlnValThrSerThrProAspGlyGlnGlnAsnPheLysThrCy 67
|||||
234 GTGTGCCAGGTAACTGACGCGCCAGCAAGCAACAGAACTTTAAACATG 283
|||||
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
284 TGCTGAGCTGGTTCCAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTTTC 333

```

```

84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
334 TGCTTGAGGCATTTGACTTCTATTGCACGGGACCTCGCAGAGAGCTACAC 383
|||||
101 LeuSerGlnProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
|||||
384 CTCTCTGAACCACTGGGTGGGAAACTTTTGAAGAATACACCCAGCTGC 433
|||||
117 aArgGluCysGlyLeuTyrLeuSerLeuGlyPheHisGluArgGlyG 134
|||||
434 CAGGGAATGTGGACTCTGGCTGTCTTGGGTGGTTCCATGAGCGTGGCC 483
|||||
134 InAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
|||||
484 AAGACTGGGAGCAGACTCAGAAAAATCTACAATGTGCACGTGCTCTCAAC 533
|||||
151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
|||||
534 AGCAAGGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT 583
|||||
167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
|||||
584 AGAGATTCCAGGCGAGGGGCTATGTGTGAAGCAACTCTACCAATGCTG 633
|||||
184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
634 GGCCCACTCTTGTAGTCACCTGTGCAGCACACCAGCAGGCAAGATTGGT 683
|||||
201 AlavalCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
|||||
684 GCTGCTCTGATGACATGCGGTTCCTTGAACCTCTCTCTGTCATTGGCT 733
|||||
217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
|||||
734 AGCTGGAGCAGAGATACTTACCTATCCTCTCAGCTTTTGGATCCATTAC 783
|||||
234 lyProAlaHisTrpGluValLeuLeuArgAlaAlaIleGluThrGln 250
|||||
784 GCCCAGCCCACTGGGAGGTGTTGTCGGGCGCGTGTATCGAAACCCAG 833
|||||
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisGluLysArgAl 267
|||||
834 TGCTATGTAGTGGCAGCAGCAGTAGTGTGGACGCCCATGAGAAGAGAGC 883
|||||
267 aSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAla 284
|||||
884 AGTTATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGGTGGCCC 933
|||||
284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgileAspLeuAsnTyr 300
|||||
934 GCTGCTCTGAGGGCCAGGCTCTGCTCTTGGCCGAATAGACCTCAACTAT 983
|||||
301 LeuArgGlnLeuArgHisLeuProValPheGlnHisArgArgProAs 317
|||||
984 CTGCGACAGTTGGCGGCACACCTGCCTGTGTTCCAGCACCGCAGGCCTGA 1033
|||||
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
1034 CCTCTATGGCAATCTGGGTCAACCCACTGTCT 1064

```

seq\_name: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq:US-09-357-675A-1

```

seq_documentation_block:
; Sequence 1, Application US/09357675A
; GENERAL INFORMATION:
; APPLICANT: Croce, M.D., Carlo M.
; TITLE OF INVENTION: Nitrilase Homologs
; FILE REFERENCE: CRO01.NP001
; CURRENT APPLICATION NUMBER: US/09/357,675A
; CURRENT FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/093,350
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 18

```

```
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 1416
/ TYPE: DNA
/ ORGANISM: Homo Sapiens
/ FEATURE:
/ NAME/KEY: gene
/ LOCATION: (0)...(0)
/ OTHER INFORMATION: n=a
US-09-357-675A-1

alignment_scores:
    Quality: 1743.00      Length: 326
    Ratio: 5.347          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-357-675C-21 x US-09-357-675A-1 ..
Align seg 1/1 to: US-09-357-675A-1 from: 1 to: 1416

2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysPr 18
|||||
111 CTGGGCTTCATCACCAGGCTCTCTCAGATTCTCTGCTCCCTTCTGTGTC 160

18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
|||||
161 TGGACTCCGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCCA 210

35 rGalaMetaIleSerSerSerCysGluLeuProLeuValAlaVal 51
|||||
211 GAGCCATGGCTATCTCTCTCTCTCTGCGAAGTCCCTGGTGCTGTG 260

52 CysGlnValThrSerThrProAspLysGlnGlnAsnPhelystThrCysAl 68
|||||
261 TGGCAGGTAACATCGAGCGCCAGACAGACAGAACTTTAAAAACATGTG 310

68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
|||||
311 TGAGCTGGTTTCGAGAGGCTGCCAGACTGGGTGCTGCTGGCTTCTGTC 360

85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
|||||
361 CTGAGGCATTGTGCTTCTCATTGCAGCGGACCTGCAGAGACGCTACACCTG 410

102 SerGluProLeuGlyGlyLysLeuLeuGluGluThrGlnLeuAlaAr 118
|||||
411 TCTGAACCACTGGGTGGGAAACTTTTGGAAAGAATACACCCAGCTTGCAG 460

118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
|||||
461 GGAATGTGGACTCTGGCTGTCTTGGGTGGTTCATGAGCTGGCCAG 510

135 spTrpGluGlnThrGlnLysIleTyAsnCysHisValLeuLeuAsnSer 151
|||||
511 ACTGGAGCAGACTCAGAAAATCTACAATTTGTCCAGCTGCTGCTGAACAGC 560

152 LysGlyAlaValValAlaThrArgLysThrHisLeuCysAspValGI 168
|||||
561 AAAGGGGACGATGTGGCCACTTACAGGAGACACATCTGTGTGACGTAGA 610

168 uileProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
|||||
611 GATTCAGGCGCAGGGGCTATGTGTGAAGCAACTCTACCATGCTGGGCG 660

185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
|||||
661 CCAGTCTTGAGTCCACCTGTCTCAGCACACCAGCAGCAAGATTGCTAGCT 710

202 ValCysTyAspMetArgPheProGluLeuSerLeuAlaAlaGlnAl 218
|||||
711 GTCTGCTATGACATGGGTTTCCCTGAACCTCTCTGGCATTTGGCTCAAGC 760

218 aGlyAlaGluIleLeuThrTyProSerAlaPheGlySerIleThrGlyP 235
|||||
761 TGGAGCAGAGATACTTACCTATCTCTTACGTTTGGATCCATTACAGGCC 810

235 roAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251
|||||
811 CAGCCCACTGGGAGGTGTTCTCGGGCCGCTATCGAAACCCAGTGC 860

252 TyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSe 268
|||||
861 TATGTAGTGGCAGCAGCAGCAGTGTGACGCCACCACATGAGAAGAGCAAG 910

268 rTyGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC 285
|||||
911 TTATGCCACACAGCATGGTGTAGACCCCTGGGGAACAGTGGTGGCCCGCT 960

285 ysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyLeu 301
|||||
961 GCTCTGAGGGGCCAGGCTCTGCTTGGCCGAATAGACCTCAACTATCTG 1010

302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLe 318
|||||
1011 CGACAGTTGCGCGCACACCTGCTGTGTTCCAGACCCGCGAGGCTGACCT 1060

318 uTyGlyAsnLeuGlyHisProLeuSer 327
|||||
1061 CTATGGCAATCTGGGTCAACCCACTGTCT 1088

seq_name: /cgn2_6/ptodata/2/pna/US093_COMB.seq:US-09-357-675C-1
seq_documentation_block:
/ Sequence 1, Application US/09357675C
/ GENERAL INFORMATION:
/ APPLICANT: Croce, Carlo M.
/ TITLE OF INVENTION: Nitrilase Homologs
/ FILE REFERENCE: CRO01.NP001
/ CURRENT APPLICATION NUMBER: US/09/357,675C
/ PRIORITY FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: 60/093,350
/ PRIOR FILING DATE: 1998-07-20
/ NUMBER OF SEQ ID NOS: 31
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 1416
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (19)...(19)
/ OTHER INFORMATION: n=a
US-09-357-675C-1

alignment_scores:
    Quality: 1743.00      Length: 326
    Ratio: 5.347          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-357-675C-21 x US-09-357-675C-1 ..
Align seg 1/1 to: US-09-357-675C-1 from: 1 to: 1416

2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysPr 18
|||||
111 CTGGGCTTCATCACCAGGCTCTCTCAGATTCTCTGCTCCCTTCTGTGTC 160

18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgProA 35
|||||
161 TGGACTCCGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCCA 210

35 rGalaMetaIleSerSerSerCysGluLeuProLeuValAlaVal 51
|||||
211 GAGCCATGGCTATCTCTCTCTCTCTGCGAAGTCCCTGGTGCTGTG 260

52 CysGlnValThrSerThrProAspLysGlnGlnAsnPhelystThrCysAl 68
|||||
261 TGGCAGGTAACATCGAGCGCCAGACAGACAGAACTTTAAAAACATGTG 310

68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85
|||||
311 TGAGCTGGTTTCGAGAGGCTGCCAGACTGGGTGCTGCTGGCTTCTGTC 360

85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101
|||||
361 CTGAGGCATTGTGCTTCTCATTGCAGCGGACCTGCAGAGACGCTACACCTG 410

102 SerGluProLeuGlyGlyLysLeuLeuGluGluThrGlnLeuAlaAr 118
|||||
411 TCTGAACCACTGGGTGGGAAACTTTTGGAAAGAATACACCCAGCTTGCAG 460

118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
|||||
461 GGAATGTGGACTCTGGCTGTCTTGGGTGGTTCATGAGCTGGCCAG 510

135 spTrpGluGlnThrGlnLysIleTyAsnCysHisValLeuLeuAsnSer 151
|||||
511 ACTGGAGCAGACTCAGAAAATCTACAATTTGTCCAGCTGCTGCTGAACAGC 560

152 LysGlyAlaValValAlaThrArgLysThrHisLeuCysAspValGI 168
|||||
561 AAAGGGGACGATGTGGCCACTTACAGGAGACACATCTGTGTGACGTAGA 610

168 uileProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
|||||
611 GATTCAGGCGCAGGGGCTATGTGTGAAGCAACTCTACCATGCTGGGCG 660

185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
|||||
661 CCAGTCTTGAGTCCACCTGTCTCAGCACACCAGCAGCAAGATTGCTAGCT 710

202 ValCysTyAspMetArgPheProGluLeuSerLeuAlaAlaGlnAl 218
|||||
711 GTCTGCTATGACATGGGTTTCCCTGAACCTCTCTGGCATTTGGCTCAAGC 760
```



```
152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValG1 168
|||||
2584 AAAGGGCAGTAGTGCCACTTACAGGAAGACACATCTGTGTGACGTAGA 2535
|||||
168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
|||||
2534 GATTCAGGGCAGGGGCTATGTGTGAAGCAACTCTACCATGGCTGGGC 2485
|||||
185 toSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
|||||
2484 CCAAGTCTTGAGTCACCTGTACGACACACAGCAGCAAGATTGGTCTAGCT 2435
|||||
202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218
|||||
2434 GTCTGCTATGACATCGGGTTCCTCAACTCTCTCGGCATTGGCTCAAGC 2385
|||||
218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
|||||
2384 TGGAGCAGAGATACTTACCTATCTTCAAGCTTTTGGATCCATTACAGGCC 2335
|||||
235 roAlaHisTrpGluValLeuLeuArgAla .ArgAlaIleGluThrGlnCy 251
|||||
2334 CAGCCCACTGGGAGGTGTGTGCGGGCCCGCTGTATCGAAACCCAGTG 2285
|||||
251 sTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlas 268
|||||
2284 CTATGTAGTGCGCAGCAGCAGCTGTGGAGCCACCACCATGAGAAGAGCAA 2235
|||||
268 eTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArg 284
|||||
2234 GTTATGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGTGTGCCCGC 2185
|||||
285 CysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrIle 301
|||||
2184 TGCTCTGAGGGCCAGGCTCTGCCCTGCCCGAATAGACCTCAACTATCT 2135
|||||
301 uArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspL 318
|||||
2134 GCGACAGTTGCGCCGACACCTGCCTGTGTTCAGCAGCAGCCGAGCCGCTGACC 2085
|||||
318 euTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
2084 TCTATGGCAATCTGGGTCAACCACACTGTCT 2056
|||||
```

seq\_name: /cgn2\_6/ptodata/2/pna/PCTUS\_COMB.seq:PCT-US00-05988-692

```
seq_documentation_block:
; Sequence 692, Application PC/TUS0005988
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Human Prostate Cancer Associated Gene Sequences and Polypeptides
; FILE REFERENCE: PA101PCT
; CURRENT APPLICATION NUMBER: PCT/US00/05988
; CURRENT FILING DATE: 2000-03-08
; EARLIER APPLICATION NUMBER: 60/124,270
; EARLIER FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 692
; LENGTH: 1382
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US00-05988-692
```

alignment\_scores:  
Quality: 1718.00 Length: 327  
Ratio: 5.270 Gaps: 0  
Percent Similarity: 99.694 Percent Identity: 99.694

alignment\_block:  
US-09-357-675C-21 x PCT-US00-05988-692

```
Align seg 1/1 to: PCT-US00-05988-692 from: 1 to: 1382
1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17
|||||
77 ATGCTGGGCTTCATCACCAGGCTCCTCACAGATTCCTGTCCCTTCTGTG 126
|||||
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
|||||
127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGC 176
|||||
34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50
|||||
177 CCAGAGCCATGGCTATCTCTCTCTCTCGAACTGCCCTGTGTGGCT 226
|||||
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67
|||||
227 GTGTGCCAGGTAAATCGACGCCAGCAGCAACAGAACTTTAAACATG 276
|||||
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
277 TCCTCAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTTC 326
|||||
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
327 TCCTTGAGGCATTTGACTTTCATTGTCACGGGACCTTCAGAGACGCTAC 376
|||||
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluTyrThrGlnLeuAl 117
|||||
377 CTGTCTGAACCACTGGGTGGGAACTTTTGGAGAATACACCCAGCTTGC 426
|||||
117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
|||||
427 CAGGGAATGTGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCGTGCC 476
|||||
134 InAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
|||||
477 AAGACTGGGACAGACTCAGAAATCTACAATGTGCAGTGTCTGTGAC 526
|||||
151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
|||||
527 AGCAAGGGCAGTAGTGCCACTTACAGGAAGACACATCTGTGTGACGT 576
|||||
167 IcluleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
|||||
577 AGAGATTCCAGGGCAGGGCT .ATGTGTGAAGCAACTCTACCATGCCTG 625
|||||
184 IyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
626 GGCCCGAGTCTTGAGTCACCTGTTCAGCAGCAGCAGGCAAGATTGCTTA 675
|||||
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217
|||||
676 GCTGTCTCTATGACATGGGTTCCTGAACTCTCTGGCATTGGCTCA 725
|||||
217 naAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234
|||||
726 AGCTGGAGCAGAGATACTTACCTATCTCTCAGCTTTTGGATCCATTAC 775
|||||
234 IyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250
|||||
776 GCCCAGCCCACTGGGAGGTGTGTGCGGGCCCGTCTATCGAAACCCAG 825
|||||
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl 267
|||||
826 TGCTATGTAGTGGCAGCAGCAGTGTGGACCCACCATGAGAGAGAGC 875
|||||
267 sSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAla 284
|||||
876 AAGTTATGCCACAGCATGGTGTGAGACCCCTGGGGAAACAGTGTGTGG 925
|||||
284 rcCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
|||||
```



926 GCTGCTCTGAGGGCCAGGCGCTTGTGCTTCCCGAATAGACCTCAACTAT 975  
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgProAs 317  
|||||  
976 CTCGACAGTTGGCCGACACCTGCTGTGTTCCAGCACCGCAGGCTGA 1025  
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327  
|||||  
1026 CCTCTATGGCAATCTGGGTCAACCCACTGTCT 1056  
seq\_name: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq:US-09-925-300-692  
seq\_documentation\_block:  
; Sequence 692, Application US/09925300  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 692  
; LENGTH: 1382  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-300-692

alignment\_scores:  
Quality: 1718.00 Length: 327  
Ratio: 5.270 Gaps: 0  
Percent Similarity: 99.694 Percent Identity: 99.694  
alignment\_block:  
US-09-357-675c-21 x US-09-925-300-692 ..  
Align seg 1/1 to: US-09-925-300-692 from: 1 to: 1382  
1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17  
|||||  
77 ATGCTGGCTTCAACACAGGCTCTCTACAGATTCTCTGCTCCCTCTGTG 126  
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34  
|||||  
127 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGC 176  
34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50  
|||||  
177 CCAGAGCCATGGGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 226  
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67  
|||||  
227 GTGTGCCAGTACATCGAGCCGACAGCAACAGCACTTTAAACATG 276  
67 salArgLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
|||||  
277 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTGCTGGCTTCC 326  
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
|||||  
327 TGCTGAGGCAATTTGACTTCTATTGACGGGACCTGCAGAGAGGCTACAC 376  
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluThrThrGlnLeuAl 117  
|||||  
377 CTGTCTGAACCACTGGGTGGGAACATTTTGGGAAGATAACACCGCTGC 426  
117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGly 134

|||||  
427 CAGGGAATGTGGACTCTGCTGTCTTGGGTGCTTCCCATGAGCGTGGCC 476  
134 InAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150  
|||||  
477 AAGACTGGGAGCAGACTCAGAAAATCTACAATTTGTACACGTGCTGAAC 526  
151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167  
|||||  
527 AGCAAGGGGCGAGTAGTGCCACTTACAGGAACACATCTGTGTGACGT 576  
167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
|||||  
577 AGAGATTCCAGGCGAGGGCT.ATGTGTGAAGCAACTCTACCATGCGCTG 625  
184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
|||||  
626 GGGCCAGTCTTGAGTCACCTGTACGACACACAGCAGCAGCAAGATTGGTCTA 675  
201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
|||||  
676 GCTGCTGTATGACATGCGGTTCCTTGAACCTCTCTTGGCATTTGGCTCA 725  
217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234  
|||||  
726 AGCTGGAGCAGAGATACTTACCTATCCTTTCAGCTTTTGGATCCATTACAG 775  
234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250  
|||||  
776 GCGCAGCCACTGGGAGGTGTTGCTGCGGCGCGTGTATCGAAACCCAG 825  
251 CysTyrValValAlaAlaAlaGlnCysGlyArgHisGlnLysArgAl 267  
|||||  
826 TGCTATGTAGTGACAGCAGCAGCTGTGGACGCCACCATGAGAAGAGAGC 875  
267 aserTyrGlyHisSerMetValValAspProTrpGlyThrValValAla 284  
|||||  
876 AAGTTATGGCCACAGCATGCTGTAGACCCCTGGGAACAGTGGTGGCC 925  
284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300  
|||||  
926 GCTGCTCTGAGGGGCCAGGCTCTGCTTCCCGAATAGACCTCAACTAT 975  
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgProAs 317  
|||||  
976 CTCGACAGTTGGCCGACACCTGCTGTGTTCCAGCACCGCAGGCTGA 1025  
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327  
|||||  
1026 CCTCTATGGCAATCTGGGTCAACCCACTGTCT 1056  
seq\_name: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq:US-60-172-373-14716  
seq\_documentation\_block:  
; Sequence 14716, Application US/60172373  
; GENERAL INFORMATION:  
; APPLICANT: Morris, MacDonald  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polym  
; FILE REFERENCE: GX-0006 P  
; CURRENT APPLICATION NUMBER: US/60/172,373  
; CURRENT FILING DATE: 1999-12-16  
; NUMBER OF SEQ ID NOS: 25,772  
; SOFTWARE: PERL Program  
; SEQ ID NO 14716  
; LENGTH: 5035  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 480310.9c



```
|||||
625 CTGAGGCAATTGGACTTCAATTCGACGGGACCTCGACAGACGCTACACCTG 674
102 SerGluProLeuGlyGlyLysLeuLeuGluThrThrGlnLeuAlaAr 118
|||||
675 TCtGAACCACTGGTGGGAACCTTTGGAGAGATACACCCAGCTGGCAG 724
118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135
|||||
725 GGAATGTGGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCGTGGCCAAG 774
135 sPrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151
|||||
775 ACTGGGAGCAGACTCAGAAATCTACAAATGTCACGTGCTGCTGAACAGC 824
152 LysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValG1 168
|||||
825 AAAGGGCAGTAGTGGCCACTTACAGAAACACACATCTGTGTGACGTAGA 874
168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185
|||||
875 GATTCCAGGGCAGGGGCTATGTGTGAAGCAACTCTACCATGCTGGGC 924
185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
|||||
925 CCAGTCTTGAGTCACCTGTGACACACACGACGAGGCAAGATTGGTCTAGCT 974
202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaGlnAl 218
|||||
975 GTCTGTATGACATCGCGTTCCTCGAACTCTCTGGCATTTGGCTCAAGC 1024
218 aGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235
|||||
1025 TGGAGCAGAGATACTTACCTATCCTTCAGCTTTTGGATTCATTACAGGCC 1074
235 roAlaHisTrp.GluValLeuLeuArgAlaArgAlaIleGluThrGlnCy 251
|||||
1075 CAGCCCACTGGGGAGGTGTGTCTGCGGGCCGCTGTATCGAAACCCAGTG 1124
251 sTyrValValAlaAlaAlaGlnCysGlyArgHisGluLysArgAlas 268
|||||
1125 CTATGTAGTGGCAGCAGACAGTGTGAGCCACCACCATGAGAAGAGACAA 1174
268 eTyrGlyHisSerMetValValAspProTrp.GlyThrValValAlaAr 284
|||||
1175 GTTATGCCACAGCATGGTGTAGACCCCTGGGGCAACAGTGTGGCCCG 1224
284 gCysSerGluClyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrL 301
|||||
1225 CTGCTCTGAGGGGCCAGGCCCTGCTGCTTGGCCGAATAGACCTCAACTATC 1274
301 euArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 317
|||||
1275 TGCACAGTGTGGCGCAGACCTGCCTGTGTCCAGACCCGAGCCCTGAC 1324
318 LeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
1325 CTCATGTGCAATCTGGGTCAACCCACTCTCT 1354
```

```
seq_name: /cqn2_6/ptodata/2/pna/us093_comb.seq:us-09-359-922-12822
seq_documentation_block:
; Sequence 12822, Application US/09359922
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS cdNA
; TITLE OF INVENTION: LIBRARIES
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 13203
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12822
; LENGTH: 3724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-12822

alignment_scores:
Quality: 1626.50      Length: 331
Ratio: 5.036          Gaps: 7
Percent Similarity: 97.583      Percent Identity: 96.677

alignment_block:
US-09-357-675C-21 x US-09-359-922-12822  ..

Align seg 1/1  to: US-09-359-922-12822  from: 1  to: 3724

4  PheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLe 20
|||||
383 TTCATCACAGGCGCTCCTCACAGATTCTGTCCCTTCTGTCTCCTGGACT 432
20  uArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaM 37
|||||
433 CCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGCCAGAGCCA 482
37  eAlaIleSerSerSerCysGluLeuProLeu.ValAlaValCysG1 53
|||||
483 TGGTATCTCTCTCTCTCTCGCAACTGCCCTGGGTGGCTGTGTGCCA 532
53  nValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAlaGluL 70
|||||
533 GGTAACATCGAGCGCCAGACAAGCAACAGAACTTTAAACATGTGCTGAGC 582
70  euValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGlu 86
|||||
583 TGGTTCAGAGGGTGCAGAGCTGGGTGCCTGCCCTGCTTCTGCTGCTGAG 632
87  AlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerG1 103
|||||
633 GCATTTTACATTCATTTGCACGGGACCCCTGCAGAGACGCTACACCTGTCTGA 682
103  uProLeu...GlyGlyLysLeuLeuGlu.GluTyrThrGlnLeuAlaArg 118
|||||
683 ACCACTGGTGGGGGAAACTTTTGGGAAGGAATACACCCAGCTTGCCAGG 732
119  GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAs 135
|||||
733 GAATGTGGACTCTGGCTGTCTTGGGTGGTTTCCATGAGCGTGGCCAAGA 782
135  pTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerL 152
|||||
783 CTGGAGCAGACTCAGAAATCTACAATTTGCACGTGCTGCTGAACAGCA 832
152  ysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGlu 168
|||||
833 AAGGGCAGTAGTGGCCACTTTACAGGAAGACACATCTGTGTGACGTAGAG 882
169  IlePro.GlyGlnGlyPro.MetCysGluSerAsnSerThrMetProGly 184
|||||
883 ATTCAGGGGCGAGGGGCTATTGTGTGAAAGCAACTCTTACCATGCTGGG 932
185  proSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAl 201
|||||
933 CCCAGCTTTGAGTCACCTGTCTCAGCACACCAGCAGGCAAGATTGGTCTAGC 982
201  aValCysTyrAspMetArgPheProGluLeuSerLeuAlaGlnA 218
|||||
983 TGTCTGTATGACATGCGGTTCCTCCCTGAACCTCTCTCTGGCATTTGGCTCAAG 1032
218  laGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGly 234
|||||
1033 CTGGAGCAGAGATACTTACCTATCTCTCAGCTTTTGGATTCATTACAGGC 1082
```

```
235 ProAlaHisTrp.GluValLeuLeuArgAlaArgAlaIleGluThrGlnC 251
|||||
1083 CCAGCCCACTGGGAGGTGTGCTGGCGGCCGCTGTATCGAAACCCAGT 1132
|||||
251 ySTyrValValAlaAlaGlnCysGlyArgHisHisGluLysArgAla 267
|||||
1133 GCTATGTAGTGGCAGCAGCAGTGTGGCCGCCACCATGAGAAGAGCA 1182
|||||
268 SerTyr.GlyHisSerMetValValAspProTrpGlyThrValValAla 284
|||||
1183 AGTTATGGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGGTGCCC 1232
|||||
284 rGCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
|||||
1233 GCTGCTCTGAGGGCCAGGCTCTGCGCTTGGCCGAATAGACCTCAACTAT 1282
|||||
301 LeuArgGlnLeuArgHisLeuProValPheGlnHisArgArgProAs 317
|||||
1283 CTGGCAGAGTTGCCCGCACACCTGCCTGTGTCCAGCACCGCAGGCCTGA 1332
|||||
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
1333 CCTCTATGGCAATCTGGGTCAACCCACTGTCT 1363
|||||
```

seq\_name: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq:US-09-359-922-12822

```
seq_documentation_block:
; Sequence 12822, Application US/09359922A
; GENERAL INFORMATION:
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Liu, Jin
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
; FILE REFERENCE: 20411-752CON1
; CURRENT APPLICATION NUMBER: US/09/359,922A
; CURRENT FILING DATE: 1999-07-22
; EARLIER APPLICATION NUMBER: US 09/205,155
; EARLIER FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: US 09/034,341
; EARLIER FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 13203
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12822
; LENGTH: 3724
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-359-922-12822
```

```
alignment_scores:
Quality: 1626.50      Length: 331
Ratio: 5.036          Gaps: 7
Percent Similarity: 97.583      Percent Identity: 96.677
```

alignment\_block:

US-09-357-675C-21 x US-09-359-922-12822

Align seg 1/1 to: US-09-359-922-12822 from: 1 to: 3724

```
4 PheIleThrArgProProHisArgPheLeuSerLeuLeuCysProGlyLe 20
|||||
383 TTCATCACCAAGGCTCTCTACAGATTCTGCTCCCTTCTGTGCTGGACT 432
|||||
20 uArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAla 37
|||||
433 CCGGATACCTCAACTCTCACTAGTCTTGTGTGTCAGCCCGCCAGAGCCA 482
|||||
37 etAlaIleSerSerSerCysGlnLeuProLeu.ValAlaValCysG1 53
|||||
483 TGGCTATCTCTCTTCTCTCTGGCAACTGCCCTGGGTGGCTGTGGCCA 532
|||||
53 nValThrSerThrProAspLysGlnGlnAsnPheLysThrCysAlaGluL 70
|||||
```

seq\_name: /cgn2\_6/ptodata/2/pna/US6036\_COMB.seq:US-60-360-207-3686

seq\_documentation\_block:

; Sequence 3686, Application US/60360207

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig

```
|||||
533 GGTAAACATCGACCCAGACAAGCAACAGAACTTTAAAAACATGTGCTGAGC 582
|||||
70 euValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGlu 86
|||||
583 TGGTTCCAGAGAGGTGCCAGACTGGGTGGCTTCCTGGCTTTCTGCTGAG 632
|||||
87 AlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerG1 103
|||||
633 GCATTTGACTTCATTTGCACGGGACCCCTGCAGAGACGCTACACCTGTCTGA 682
|||||
103 uProLeu...GlyGlyLysLeuLeuGlu.GluTyrThrGlnLeuAlaArg 118
|||||
683 ACCACTGGGTGGGAAACATTTTGGGAAGCAATACACCCAGCTTGCACAG 732
|||||
119 GluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAs 135
|||||
733 GAATGTGGACTCTGGGTGCTCTTGGGTGGTTTCCATGAGCGTGCCCAAGA 782
|||||
135 pTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSerL 152
|||||
783 CTGGGAGCAGACTCAGAAAATCTACAAATTGTCAGTGTCTGCTGACGTAG 882
|||||
152 ySGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspValGlu 168
|||||
833 AAGGGCAGTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGTAGAG 882
|||||
169 IlePro.GlyGlnGlyPro.MetCysGluSerAsnSerThrMetProGly 184
|||||
883 ATTCCAGGGGAGGGGCTATTGTGTGAAAGCAACTCTACCATGGCTGGG 932
|||||
185 ProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAl 201
|||||
933 CCCAGCTCTTGAGTCACTGTACACACACCCAGCAGGCAAGATTGGTCTAG 982
|||||
201 aValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnA 218
|||||
983 TGTCTGCTATGACATGCGGTTCCTGAACCTCTCTCTGGCATTTGGCTCAAG 1032
|||||
218 laGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGly 234
|||||
1033 CTGGAGCAGAGATACTTACCTATCTTCACTTTTGGATCCATTACAGGC 1082
|||||
235 ProAlaHisTrp.GluValLeuLeuArgAlaArgAlaIleGluThrGlnC 251
|||||
1083 CCAGCCCACTGGGAGGTGTGCTGGCGGCCGCTGTATCGAAACCCAGT 1132
|||||
251 ySTyrValValAlaAlaGlnCysGlyArgHisHisGluLysArgAla 267
|||||
1133 GCTATGTAGTGGCAGCAGCAGTGTGGCCGCCACCATGAGAAGAGCA 1182
|||||
268 SerTyr.GlyHisSerMetValValAspProTrpGlyThrValValAla 284
|||||
1183 AGTTATGGGCCACAGCATGTGTGTAGACCCCTGGGGAACAGTGGTGCCC 1232
|||||
284 rGCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300
|||||
1233 GCTGCTCTGAGGGCCAGGCTCTGCGCTTGGCCGAATAGACCTCAACTAT 1282
|||||
301 LeuArgGlnLeuArgHisLeuProValPheGlnHisArgArgProAs 317
|||||
1283 CTGGCAGAGTTGGCCGACACCTGCCTGTGTCCAGCACCGCAGGCCTGA 1332
|||||
317 pLeuTyrGlyAsnLeuGlyHisProLeuSer 327
|||||
1333 CCTCTATGGCAATCTGGGTCAACCCACTGTCT 1363
|||||
```

; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF  
; FILE REFERENCE: CL001321  
; CURRENT APPLICATION NUMBER: US/60/360,207  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 47235  
; SEQ ID NO 3686  
; LENGTH: 1296  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-60-360-207-3686

alignment\_scores:  
Quality: 1487.00 Length: 327  
Ratio: 4.828 Gaps: 2  
Percent Similarity: 94.190 Percent Identity: 84.404  
alignment\_block:  
US-09-357-675C-21 x US-60-360-207-3686 ..  
Align seg 1/1 to: US-60-360-207-3686 from: 1 to: 1296

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17  
|||||  
58 ATGCTGGGCTTCATCACCAGGCTCTCTACCAA.....CTCCTGTG 98  
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgp 34  
|||||  
99 TACCGGATACCGATTACTTCGATCCAGTACTTGTACTACGCCAGGC 148  
34 roArgAlaMetAlaIleSerSerSerCysGlnLeuProLeuValAla 50  
|||||  
149 CCAGAACCACTGCC...TCATCAACTTCTGGCAGCTGCCCTGTGGCT 195  
51 ValCysGlnValThrSerThrProAspLysGlnAsnPhelyThrCy 67  
|||||  
196 GTGTGCAGGTAACATCAACACCAACCAAGCAAGAACTTTAAACATG 245  
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
|||||  
246 TGCTGAGTTGGTTCAGAGGCTCCAGACTGGGTGCTTGCCTGGCCTTC 295  
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
|||||  
296 TGCTGAGGCACTTGAATTTATTCAGCAAAACCTGCCGAGACATTACTC 345  
101 LeuSerGluProLeuGlyLysLeuLeuGluGluThrThrGlnLeuAl 117  
|||||  
346 CTGTCCGAACCACTGAATGGGGTCTTTTGGGCCAATATAGCCAGCTTC 395  
117 aArgGlyCysGlyLeuTriLeuSerLeuGlyGlyPheHisGluArgGly 134  
|||||  
396 CAGGGAATGTGGATCTGGCTGTCTTGGCGGTTTCCAGCAGCGTGGCC 445  
134 lnAspTrpGluGlnThrGlnLysIleTyAsnCysHisValLeuLeuAsn 150  
|||||  
446 AAGACTGGGAGCAGAAATCAGAAAATCATAATGTCATGTGCTTTTGAAC 495  
151 SerLysGlyAlaValValAlaThrThrArgLysThrHisLeuCysAspVa 167  
|||||  
496 AGCAAGGATCACTAGTGGCAGTTTACAGAGACACATCTGTGCGATGT 545  
167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
|||||  
546 AGAGATCCCAAGTCAGGGGCGATGAGAGAAACCACTATACCAAGCCTG 595  
184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
|||||  
596 GAGGCACTCTTGGACCACTGTCAAGACACCGCGCTGCAAGGTTGCTTA 645  
201 AlaValCysTyAspMetArgPheProGluLeuSerLeuAlaGlu 217  
|||||  
646 GCAATCTGTTATGACATGCGGTTTCCCTGAACTTTCTTTGAAATTTGGCTCA 695

217 nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG 234  
|||||  
696 AGCTGGGCGAGAAATACTTACTTATCTTCCTTCAGCTTTGGATCTGTACAG 745  
234 lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln 250  
|||||  
746 GTCAGCCCACTGGGAGGTGCTGCTGGGCGCGCCCATTTGAATCTCAG 795  
251 CysTyTrValValAlaAlaAlaGlnCysGlyArgHisGlnLysArgAl 267  
|||||  
796 TGCTATGTAAATAGCAGCAGCGAGTGTGGACGCCCATCAAAACAGAGC 845  
267 aserTyTrGlyHisSerMetValValAspProTrpGlyThrValValAla 284  
|||||  
846 AACTTATGGCCATAGCATGTGTGTGACCCGTGGGCGACAGTGGTGCC 895  
284 rgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300  
|||||  
896 GCTGCTCCGAGGACCGAGGCTCTGCTTGTGCTGCAATTTGATCTCCACTT 945  
301 LeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAs 317  
|||||  
946 CTACAACAGATGCCCAACACCTGCTGTGTTCCAGCAGCCGACACCTGA 995  
317 pLeuTyTrGlyAsnLeuGlyHisProLeuSer 327  
|||||  
996 CCTCTATGGCAGTCTGGGTCTCATCTCTCT 1026  
seq\_name: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq:us-60-242-679-78  
seq\_documentation\_block:  
; Sequence 78, Application US/60242679  
; GENERAL INFORMATION:  
; APPLICANT: Ladunga, Steven Istvan  
; APPLICANT: Spier, Eugene  
; APPLICANT: Greenberg, Simon  
; APPLICANT: Brandenberger, Ralph  
; APPLICANT: Wang, Yu  
; APPLICANT: Dubman, Alex  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; FILE REFERENCE: CL000898-PROV  
; CURRENT APPLICATION NUMBER: US/60/242,679  
; CURRENT FILING DATE: 2000-10-24  
; NUMBER OF SEQ ID NOS: 2265  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 78  
; LENGTH: 3620  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(3620)  
; OTHER INFORMATION: n = A,T,C or G  
US-60-242-679-78

alignment\_scores:  
Quality: 1468.50 Length: 660  
Ratio: 4.505 Gaps: 8  
Percent Similarity: 49.394 Percent Identity: 48.939  
alignment\_block:  
US-09-357-675C-21 x US-60-242-679-78/rev ..  
Align seg 1/1 to reverse of: US-60-242-679-78 from: 1 to: 3620  
2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 18  
|||||  
2918 CTGGGTTTCATCACCAGGCTCTCTCAGATTCCTGTCCTTCTGTGTCC 2869

18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnPro..... 32  
|||||  
2868 TGGACTCCGGATACCTCAACTCTCAGTACTTTTGTGCTCAGCCAGGTAAC 2819  
32 ..... 32  
2818 ACCTTTTGTGTCTCAGTGCCTGCCACTTAGATGCTCAGTTGTAA 2769  
32 ..... 32  
2768 ATGGATAGTGGGAGACACAGGAGGTCAACTATCCACACATTTGATGGT 2719  
32 ..... 32  
2718 GAGCCCTACTAGCCCTGGGTCAACGTGCCCTGTAGAGCATGATCAAAAG 2669  
32 ..... 32  
2668 GAAGTCCAGCTTTCTCCTCCTCCACTTCCACTTGCACCCCTTAGCATTAATTTG 2619  
33 .....ArgPr 34  
2618 CTTCCCTGTGCTATGAATCTGAGAATCTGCTATGCTGTTCACAGGCC 2569  
34 oArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAlaVal 51  
|||||  
2568 CAGAGCCATGGCTATCTCCTCTTCTCCTCGCAACTGCCCTGGTGGCTG 2519  
51 alCysGlnValThrSerThrProAspLysGlnGlnAsnPhelLysThrCys 67  
|||||  
2518 TGTGCCAGGTAACTCGAGCCGACAGAACGACAGAACTTTAAACATGT 2469  
68 AlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLe 84  
|||||  
2468 GCTGAGCTGGTTCGAGAGGCTGCACACTGGGGTGGCTGGCTGGCTTCCT 2419  
84 uProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisL 101  
|||||  
2418 GCCTGAGGCATTTGACTTCATTCCACGGGACCTGCAGAGACGCTACACC 2369  
101 euSerGluProLeuGlyGlyLysLeuLeuGluGluThrGlnLeuAla 117  
|||||  
2368 TGTCTGAACCACTGGGTGGGAACTTTTGGAGAAATACACCCAGCTTGCC 2319  
117 ..... 117  
2318 AGGTATCAGGGAATAGCGAGGAGAGGTAGAATCTTTGTTGGACAGTGT 2269  
117 ..... 117  
2268 CCCTGGGTTGCCAGATATGAGGCTAGAGCCTTGAGAAGTCAGTGAAGAGT 2219  
118 .....ArgGluCysGlyLeuTrpLeuSerLeuGly 127  
|||||  
2218 TGTGAGTGTCCCTTCCCCAGGGAATGTGGACTCTGGCTGTCTGGGT 2169  
128 GlyPheHisGluArgGlnAspTrpGluGlnThrGlnLysIleTyraS 144  
|||||  
2168 GGTTCATGAGGCTGGCCAAAGCTGGGAGCACTCAGAAAATCTACAA 2119  
144 nCysHisValLeuLeuAsnSer...Lys..... 152  
|||||  
2118 TTGTCAGTGTCTCTGAACAGCAAGGTGAGACTTTATAACCCCTTAGCC 2069  
152 ..... 152  
2068 TGCCTCTTCCCATGCTCTTCTACCTAGATTCTCCAGATTGTTCTCAAC 2019  
152 ..... 152  
2018 TCTATTTCCTGACCAAGGATTTAGGGGTGGTCTCTACTTCACTTCTCCTA 1969  
152 ..... 152

1968 GCCTATAAACTATCTCTCTTGGGAGGAGTAAGCAAGGCTTCTAGAACA 1919  
153 .....GlyAlaValValAlaThrT 159  
|||||  
1918 CCAGCACTGATATTCCTTCTTCTTACTGTAGGGCAGTAGTGGCCACTT 1869  
159 yArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyProMet 175  
|||||  
1868 ACAGGAAGACACATCTGTGTGACGTAGAGATCCAGGGCAGGGGCTATG 1819  
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValse 192  
|||||  
1818 TGTGAAGCAACTCTACCATGCTGCGCCACGCTTGTGAGTCACCTGTCAG 1769  
192 rThrProAlaGlyLysIleGly...Leu..... 200  
|||||  
1768 CACACCAGCAGGCAAGTAGGAGTTGTGAAGGATGAGGGAGGGAACAG 1719  
200 ..... 200  
1718 GAATACTTTGAACTGCCAGTAGAGGTAGAAAGCCCTAAGAGAGGGGTA 1669  
200 ..... 200  
1668 ATGGAAATATGACTAGATGCTGTGACAAACAGAGCAGGAAGACTACTAAG 1619  
201 .....AlaValCysTyrAspMetArg 207  
|||||  
1618 TAGGCTGTTTTTCAATCCAGATTGGTCTAGCTGTCTGTATGACATCGG 1569  
208 PheProGluLeuSerLeuAlaGlnAlaGlyAlaGluIleLeuTh 224  
|||||  
1568 TTCCCTGAACTCTCTCGCATTTGGCTCAAGCTGGAGCAGAGATCTTAC 1519  
224 rTyProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal. 240  
|||||  
1518 CTATCTTCAGCTTTTGGATCCATTACAGCCCGAGCCCTGGGAGGTAA 1469  
241 .Leu..... 241  
1468 GATGATGSCCTTTTAAACATAGAGGSCCTTTTCTTAACCTCATCTTCCCC 1419  
241 ..... 241  
1418 CCTTGGCCCTACCAGTTAAATCTTCCCTTCCCTTCCACCTAATGGGAAAC 1369  
241 ..... 241  
1368 TCATTCGCCAGATATTTCTCATGAATAGTTAAATAGTAAATCATTC 1319  
241 ..... 241  
1318 TAGGCAATTAACAAAATAGTCACAAATGGGTAGATTGGTCTGTAATGTCC 1269  
241 ..... 241  
1268 TCACCTGTCACTTCCCACTATTTGCTACATGCTACTTAAGTGAACACACA 1219  
242 .....LeuArgAlaArgAlaIleGluThrGlnCysty 252  
|||||  
1218 TCTCATGCCAGGTGTGCTGCGGGCCGCTGTATCGAAAACCCAGTGCTA 1169  
252 rValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSerT 269  
|||||  
1168 TGTAGTGGCAGCAGCAGCTGTGGAGCCCACTAGAGAGAGAGCAAGTT 1119  
269 yGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285  
|||||  
1118 ATGGCCACAGCATGTGTGAGACCCCTGGGGAACAGTGGTGGCCCGCTGC 1069  
286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeuAr 302  
|||||

1068 TCTGAGGGGCCAGGCTCTGCTTGGCCGAAATAGACCTCAACTATCTGGC 1019  
302 gGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeu 319  
1018 ACAGTTGGCGGACACCTGCTGTGTCCAGACCCGACGCGCTGACCTCT 969  
319 yGlyAsnLeuGlyHisProLeuSer 327  
968 ATGCAATCTGGGTACCCACTGCT 943  
seq\_name: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq:us-60-324-185-10674

seq\_documentation\_block:  
; Sequence 10674, Application us/60324185  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Lal, Preeti  
; APPLICANT: Diep, Dinh  
; TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING  
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE  
; FILE REFERENCE: GX-0019-1 P  
; CURRENT APPLICATION NUMBER: US/60/324,185  
; CURRENT FILING DATE: 2001-09-21  
; NUMBER OF SEQ ID NOS: 35862  
; SOFTWARE: PERL Program  
; SEQ ID NO 10674  
; LENGTH: 5482  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 1398463.12  
; NAME/KEY: unsure  
; LOCATION: 1652, 4633, 5437  
; OTHER INFORMATION: a, t, c, g, or other  
US-60-324-185-10674

alignment\_scores:  
Quality: 1315.00 Length: 354  
Ratio: 4.242 Gaps: 25  
Percent Similarity: 87.571 Percent Identity: 86.441

alignment\_block:

US-09-357-675C-21 x US-60-324-185-10674

Align seg 1/1 to: US-60-324-185-10674 from: 1 to: 5482

2 LeuGlyPheLeuThrArg.ProProHisArgPheLeuSerLeuLeuCysp 18  
680 CTGGGCTTCATCACCGAGGGCTCCTACAGATTCTCTCCCTTCTGTGTC 729  
18 roGlyLeuArgIleProGlnLeuSerValLeuCys.AlaGlnProArgPr 34  
730 CTGGGCTCCGGATACCTCAACTCTCAGTACTTTGTTGCTCAGCCCGAGGCC 779  
34 oArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAlav 51  
780 CAGAGCCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 829  
51 alCysGln.ValThrSerThrProAspPlyGlnGlnAsnPheLysThrCy 67  
830 TGTGCCAGGGTAACATCGAGCGCCAGACAAGCAAGCAAGCAAGCAAGCAATG 879  
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
880 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTTTC 929  
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
930 TGCCTGAGGCATTTGACTTCTATTCACGCGGACCTGCAGAGACGCTACAC 979

101 LeuSerGluProLeuGly.GlyLysLeuLeuGluGluTyThrGlnLeuA 117  
980 CTCTCTGACCACTGGGTGGAACTTTTGGGAAGATACACCCAGCTTG 1029  
117 laArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArg.Gl 133  
1030 CCAGGGAATGTGGACTCTGGCTGTCTTGGGTGGTTCATGAGCGTGGG 1079  
133 yGlnAspTrpGluGlnThrGlnLysIleTyAsnCysHisValLeuLeuA 150  
1080 CCAAGACTGGGAGACACTCAGAAATCTACAATGTTCACGTGCTGCTGA 1129  
150 snSerLysGlyAlaValVal.AlaThr...TyArgLysThrHisLeuCy 165  
1130 ACAGCAAGGGGAGTAGTGGGCACTTAACAGGGAAGGACACATCTGTG 1179  
165 sAsp.ValGluIlePro.GlyGlnGlyProMetCysGluSerAsnSerTh 181  
1180 TGAACGTAGAGATTCCAAGGGCAGGGGCTTATGTGAAAGCAACTCTAC 1229  
181 rMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 198  
1230 CATGCCCTGGCCAGCTTTCAGTCACCTGTTCAGCACACACAGCAGCA 1279  
198 leGlyLeuAlaValCysTyAspMetArgPheProGluLeuSerLeuAla 214  
1280 TTGGTCTAGCTGTCTGTATGACATGGGTTCCTGAACTCTCTCTGCA 1329  
215 LeuAlaGlnAlaGlyAlaGluIleLeuThrTyProSerAlaPheGlySe 231  
1330 TTGGCTCAAGCTGGAGCAGAGATACCTTACCTATCTCTCAGCTTTTGGATC 1379  
231 rIleThrGlyProAlaHis.TrpGlu.ValLeuLeu.ArgAlaArgAlaI 247  
1380 CATACAGGCCAGCCCACTTGGGAAGGTGTGCTGGCGGGCC..... 1422  
247 leGluThrGlnCysTyr.....ValValAlaAla 256  
1423 .....CCGTGCTATCCGAAACCCAGTCGCTTATGTAAGTTGGCAGCA 1464  
257 AlaGlnCys.GlyArg.HisHisGluLysArgAlaSerTyThrGlyHisSer 272  
1465 GCACAGTGTGGGACGGCCACCACATGAGAAAGAGCAAGTTATGGCCACAG 1514  
273 Met.ValValAspProTrp.GlyThrValValAlaArgCysSerGlu.Gl 288  
1515 ATGGGTGGTAGACCCCTGGGGGACAGTGTGGCCCTCTCTCTAGGG 1564  
288 yProGlyLeuCys.LeuAlaArgIleAspLeuAsnTyThrLeuArgGlnLeu 304  
1565 GCCAGGCTCTGCCCTTGGCCGAATAGACCTCAACTATCTGCGACAAGTT 1614  
305 ArgArgHis...LeuProVal.PheGlnHisArgArg.ProAsp.LeuTy 319  
1615 GCOCCTGACACCCCTGCTGTGTTCAGCACCCGAGGNCCTGACCCCTTA 1664  
319 rGlyAsn.LeuGlyHisPro 325  
1665 TGGCAATCTTGGGTACCC 1684





OM of: US-09-357-675C-21 to: Pending\_Patents\_NA\_New:\* out\_format : pfs

Date: Apr 29, 2002 12:20 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=Cgn2\_1/USPTO.spool/US09357675/runat\_29042002\_091758\_3352/app\_query.fasta\_1.1208  
-DB=Pending\_Patents\_NA\_New -QFWT=fastap -SUFFIX=p2n.rnpn  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -GGAPOP=4.500 -GGAPEXT=0.050 -YGAPOP=10.000  
-YGAPEXT=0.500 -YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675 -CGN1\_1\_88 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-357-675C-21

Query length: 327

Database: Pending\_Patents\_NA\_New:\*

Database sequences: 139564

Database length: 91993600

Search time (sec): 106.130000

score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq	US-60-365-384-491	491.50	705.56	4.8e-32	1
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-209B-235	259.50	370.09	2.3e-13	1
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-210B-33512	233.00	341.77	8.8e-12	1
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-210B-19553	189.50	280.31	2.3e-08	1
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-209B-466	185.00	255.76	5.4e-07	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-102-524-748	129.50	191.56	0.0020	2
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-209B-1148	126.50	177.18	0.0129	1
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-210B-5721	124.00	184.19	0.0053	1
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-210B-5426	118.50	176.78	0.0136	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-115-123-90	95.00	128.46	6.67	10
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-14411	93.00	102.16	194.66	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-14785	93.00	102.16	194.66	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-108	92.50	121.10	17.15	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-103-299-1860	92.50	118.13	25.10	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-103-295-384	92.50	96.83	385.55	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-103-295-385	92.50	96.61	396.24	2
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-540-210B-35282	91.50	135.12	2.84	1
/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq	US-60-365-384-61	91.50	118.80	23.01	16
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-109-528-1	91.00	110.19	69.46	385
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-113-123-126	89.50	118.53	23.84	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-108-605-74	89.50	114.43	40.34	19
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-09-413-186A-3	89.50	108.23	89.38	37
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-10831	89.00	103.71	159.51	7
/cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq	US-60-365-384-472	88.50	100.88	229.27	7
/cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq	US-09-627-465B-2	87.50	113.52	45.31	15
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-118-328-1	87.50	110.69	65.16	213
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-14118	87.50	82.18	2.5e+03	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-09-975-254-8983	87.00	130.91	4.87	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-107-806-80	87.00	106.11	117.20	32
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-108-605-210	87.00	100.22	249.51	5
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-108-605-212	87.00	99.74	285.21	8
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-8734	87.00	99.00	291.95	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-9557	86.50	104.25	148.81	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-7442	86.50	99.98	257.33	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-9491	86.50	97.67	346.14	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-107-431-280	86.50	79.72	3.4e+03	4
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-6571	86.00	113.91	43.11	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-12023	86.00	100.11	253.10	1
/cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq	US-10-105-299-12021	86.00	97.75	342.54	1

/cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:US-10-105-299-12019 + 86.00 97.11 371.91  
/cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:US-10-105-299-11976 - 86.00 91.99 716.98  
/cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:US-10-105-299-15008 - 86.00 90.03 922.16  
/cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:US-10-105-299-15009 - 86.00 82.32 2.5e+0  
/cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:US-10-105-299-11975 - 86.00 78.17 4.2e+0

seq\_name: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq:US-60-365-384-491

seq\_documentation\_block:  
; Sequence 491, Application US/60365384

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Wang, Jian-rui

; APPLICANT: Wang, Zhiwei

; APPLICANT: Zhang, Jie

; APPLICANT: Zhou, Ping

; APPLICANT: Wehrman, Tom

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Ghosh, Malabika

; APPLICANT: Zhao, Qing A.

; APPLICANT: Asundi, Vinod

; APPLICANT: Ren, Feiyan

; APPLICANT: Xue, Aidong J.

; APPLICANT: Ma, Yungqing

; APPLICANT: Wang, Dunrui

; APPLICANT: Weng, Gezhi

; TITLE OF INVENTION: Novel Nucleic Acids and

; FILE REFERENCE: 814

; CURRENT APPLICATION NUMBER: US/60/365,384

; CURRENT FILING DATE: 2002-03-14

; NUMBER OF SEQ ID NOS: 666

; SOFTWARE: pt\_FL\_genes Version 6.0

; SEQ ID NO 491

; LENGTH: 1251

; TYPE: DNA

; ORGANISM: Homo sapiens

US-60-365-384-491

alignment\_scores:

Quality: 491.50 Length: 312

Ratio: 2.398 Gaps: 10

Percent Similarity: 65.705 Percent Identity: 36.218

alignment\_block:

US-09-357-675C-21 x US-60-365-384-491/rev ..

Align seg 1/1 to reverse of: US-60-365-384-491 from: 1 to: 1251

23 ProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaI 39

1201 CCGAGGTGGTGTGTCTGC.....AGAGTCATG..... 1172

39 eSerSerSerCysGlnValLeuValAlaValCysGlnValThrS 56

1171 .....ACCTCTTCGGCTGGCCCTCATCCAGCTT...CAGATTCTT 1132

56 erThrProAspLysGlnGlnAsnPhelYthrCysAlaGluLeuValArg 72

1131 CCATC.....AAATCAGATAAGCTACTCGCGCTTAGCTTCATCCGG 1088

73 GluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAs 89

1087 GAGCAGCACAGGAGGAGCAAAATAGTTCTTTCGCCGAATGCTTT... 1040

89 pHeIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuG 106

1039 .....AATTCTCCATATG 1027

106 lyGlyLysLeuLeuGluGluTyr..... 113

1026 GAGCGAAATATTTTCCTGAATATGCAGAGAAATTCCTGTTGAATCCACA 977

```
114 .....ThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuG1 127
      : : : : : : : : : : : : : : : : : : : : : : : :
976 CAGAAGCTTCTCAAGTAGCAAGAAATGACAGCATATATCTCATTTGGAGG 927

127 yGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrA 144
      : : : : : : : : : : : : : : : : : : : : : : : :
926 CTCATCCCTGAA.....GAGGATGCTGGCAAAATATATATA 892

144 snCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArg 160
      : : : : : : : : : : : : : : : : : : : : : : : :
891 ACACCTGTGCTGTGTGGCCCTGATGGAACCTTACTAGCAAGATATAGA 842

161 LysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysG1 177
      : : : : : : : : : : : : : : : : : : : : : : : :
841 AAGATCCATCTGTTTGACATGTATGTTCTCGGAAAAATATACATTTCAAGA 792

177 uSerAsnSerThrMetProGlyProSerLeuGluSerProValSerThrP 194
      : : : : : : : : : : : : : : : : : : : : : : : :
791 ATCTAAACATTCAGTCCGGGTGATAGTTTC...TCCACATTTGATACTC 745

194 roAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGlu 210
      : : : : : : : : : : : : : : : : : : : : : : : :
744 CTTACTGCAGAGTGGGTCTGGGCATCTGCTTACGACATGCGGTTTGACAG 695

211 LeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSe 227
      : : : : : : : : : : : : : : : : : : : : : : : :
694 CTTGCACAATCTACGCACAGAGAGCTGCCAGCTGTGGTATATCCAGG 645

227 rAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgA 244
      : : : : : : : : : : : : : : : : : : : : : : : :
644 AGCTTTTAATCTGACCACCTGGACCAGCCCATTTGGGAGTTACTTCAGCGAA 595

244 laArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnCysGly 260
      : : : : : : : : : : : : : : : : : : : : : : : :
594 GCGGGCTGTGTATATCAGGTATATGTGGCCACAGCCCTCTCCTGCCC... 548

261 ArgHisHisGluLysArgAlaSerTyr.....GlyHisSerMetVa 274
      : : : : : : : : : : : : : : : : : : : : : : : :
547 .....CGGGATGACAAAGCCTCTATGTTCCTGGGGACACAGCAGCT 504

274 lValAspProTrpGlyThrValValAlaArgCysSerGluGlyProGlyL 291
      : : : : : : : : : : : : : : : : : : : : : : : :
503 GGTGAACCCCTTGGGGGAGGTTCTAGCCAAAGCTGGCAGAGAAGACAA 454

291 euCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArgArgHis 307
      : : : : : : : : : : : : : : : : : : : : : : : :
453 TCGTGATTTCAGACATAGACCTGAAGAAGCTGGCTGAATACCCAGCAA 404

308 LeuProValPheGlnHisArgArgProAspLeuTyr 319
      : : : : : : : : : : : : : : : : : : : : : : : :
403 ATCCCCGTTTTAGACAGAGCGCATCAGACCTCTAT 368
```

seq\_name: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:US-09-540-209B-235

```
seq_documentation_block:
; Sequence 235, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 235
; LENGTH: 888
; TYPE: DNA
; ORGANISM: B.fragilis
US-09-540-209B-235

alignment_scores:
```

```
Quality: 259.50 Length: 298
Ratio: 1.612 Gaps: 8
Percent Similarity: 54.027 Percent Identity: 28.188

alignment_block:
US-09-357-675C-21 x US-09-540-209B-235 ..
Align seg 1/1 to: US-09-540-209B-235 from: 1 to: 888

49 ValAlaValCysGlnValThrSerThrProAspLysGlnGlnAsnPheL 65
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
19 GTAGGAATCATTTCAACAGGCTAACACATCATGATATAGGATAAACCTGAT 68
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
65 sThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuA 82
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
69 GAACCTGGCTAAAGATATTGAAGCATGTGCCGCTAATGGCGCTCACCTTG 118
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
82 laPheLeuProGluAlaPheAsp.....PheIleAlaArgAspPro 95
   ||| ||| ||| :||| :||| :||| :||| :||| :||| :|||
119 TTGTTCTGCAAGAACTTCATAATTCTTTCTATTCTCTGTCAGACAGAGAAT 168
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
96 AlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluL 112
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
169 ACGGATTTATTGACTGGCAGAACCCATTCTCTGGCCCTTCTACCGGATT 218
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
112 uTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyP 129
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
219 CTATTCGGAACCTGGCGAGCAATCGGATAGTGTGTACTTCTTTGT 268
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
129 heHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCys 145
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
269 TTGAGAAACGTGCTCGGGA.....CTATATCATATAATACA 303
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
146 HisValLeuLeuAsnSerLysGlyAlaValAlaThrTyrArgLysTh 162
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
304 GCTGTTGTTCTTACCGGGATGGAAGTATTGGCGGAAATATATCGTAAGAT 353
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
162 rHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerA 179
   ||| ||| ||| :||| :||| :||| :||| :||| :||| :|||
354 GCAT.....ATCTCTGATGATCCGGCTTATTACGAGAAAT 388
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
179 snSerThrMetProGlyProSerLeuGluSerProValSerThrProAla 195
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
389 TCTATTTTAACTCCGGGAGATATTGGCTTTGAACCGATTTCAGACCTCTTTA 438
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
196 GlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSe 212
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
439 GGCAGATTGGGTGTGTTGTTGCTGGGATCAATGTTATCCGGGAAGCTGC 488
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
212 rLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlap 229
   : |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
489 TCGCCTGATGGCGTTCAAGAGGAGCTGAGATTTTGATTTATCCTACTGCTA 538
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
229 heGlySerIleThrGlyProAlaHis..... 237
   : ||| :||| :||| :||| :||| :||| :||| :|||
539 TCGGTTGGGAGACTACAGATACAGATCAGCAAAAGAAACGTCAGCTCAAT 588
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
238 ...TrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVa 253
   ||| :||| :|||:||||: ||| :|||: |||:||||: |||
589 GGTGTGATTTTCTCAGCTCGCATGCGGTAGCAATGGCGTCCCGGT 638
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
253 lValAlaAlaAlaGlnCysGlyArgHisGluLysArgAlaSer.... 268
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
639 GATTTCAGTCAATCGTGTGCGT.....CAGCAACCTGATCCCTCAGGAC 682
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
269 .....TyrGlyHisSerMetValValAspProTrp 278
   : |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
683 AGACCAACGGGATATTATTTTGGGAAATAGTTTGTTCGCGGACCGCAG 732
   |||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
279 GlyThrValValAlaArgCys.....SerGluGlyProGlyLe 291
   ||| :|||:||||: ||| :|||:||||: ||| :|||: |||:||||: |||
733 GGTGAATACCTGGCTCAGCGGGAAATGACCGCTCTGAAAT..... 774
```

[illegible]

```
|||||
142 GCCTGGGACACAGCAGCTGTGTGGATCTCTGGGGCAGGCTCTCAACCA 191
|||||
284 qCysSerGluGlyProGlyLeuGlyLeuAlaArgIleAspLeuAsnTyrL 301
|||||
192 AGCTGGCAGCGAAGAACAACTGTCTACTCAGACATAGACCTGAAGAAGC 241
|||||
301 euArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAsp 317
|||||
242 TGCTGAATTCGGCAGCAATTCCTCCATTTAAACAGAAACGACGAGAC 291
|||||
318 LeuTyr 319
|||||
292 CTCTAT 297
```

seq\_name: /sgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:US-09-540-210B-19553

```
seq_documentation_block:
; Sequence 19553, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997
; PRIOR APPLICATION NUMBER: 60/021,275
; PRIOR FILING DATE: June 25, 1995
; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
```

```
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 3654
; SOFTWARE: PERL Program
; SEQ ID NO: 19553
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00309797
; NAME/KEY: unsure
; LOCATION: 224
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-19553

alignment_scores:
Quality: 189.50 Length: 75
Ratio: 3.384 Gaps: 1
Percent Similarity: 74.667 Percent Identity: 49.333

alignment_block:
US-09-357-675C-21 x US-09-540-210B-19553
Align seg 1/1 to: US-09-540-210B-19553 from: 1 to: 239

169 IleProGlycInGlyProMetCysGluSerAsnSerThrMetProGlyPr 185
:::|||||::: ::: |||||:::
1 GTTCTGGAAAAATTACATTTCAGAAATCTAAACATCTAGTCCGGGTGA 50
```

```
185 oSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAlav 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 TAGTTC...TCCACATTTGATACCTCTTACTGCAGAGTGGTCTGGCA 97

202 alCysTyrAspMetArgPheProGluLeuSerLeuAlaGlnAla 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 TCGCTACGACATCGCGTTTGCAGAGCTTCACAAATCTACGCACAGAGA 147

219 GlyAlaGluLeuThrTyrProSerAlaPheGlySerIleThrGlyPr 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 GCGTGCAGCTGTGGTATCCAGAGCTTTAATCTGACCACCTGACC 197

235 oAlaHisThrGluValLeuLeuArg 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 AGCCATTGGGAGTTACTTCCAGCGA 222
```

seq\_name: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:US-09-540-209B-466

```
seq_documentation_block:
; Sequence 466, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 466
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: B. fragilis
US-09-540-209B-466
```

```
alignment_scores:
  Quality: 185.00      Length: 249
  Ratio: 1.350        Gaps: 11
  Percent Similarity: 55.020  Percent Identity: 28.112

alignment_block:
US-09-357-675C-21 x US-09-540-209B-466 ..
Align seg 1/1 to: US-09-540-209B-466 from: 1 to: 1599

85 ProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLe 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
918 CCGCGGCTGGCCCAATATACCGAAGAAATCCG...CGATCGCTTCATTA 964

101 uSerGluProLeuGlyLysLeuGluTyrThrGlnLeuAlaA 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
965 ATCT.....GGCAATCAGTACACATCAATAT..... 992

118 rGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGln 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
993 .....CATCACAGGAAGTAT 1007

135 AspTrpGluGlnThrGlnLysIle...TyrAsnCysHisValLeuLeuAs 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1008 GCGCTTGATCAAAAGACAGCGATGCTGTGTAACATGCCGGATTTCTTGGC 1057

150 nSerLys.GlyAlaValAlaThrTyrArgLysThrHisLeuCysAsp 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1058 GAGCGGACGA...ACTTAGAATGTACGAAAGCTCCATGTACCCCG 1104

167 ValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPr 183.
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1105 GACGAGATAAGAGTTGGGA.....CTGAG 1130

183 oGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyL 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1131 CGCGGCAAAACAGCTTAAACATTCGATACGAGTGTGCAAGATAGGCA 1180
```

```
200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1181 TACTGATCTGTTATGATGTGAATTTCCGGAACCTCTCCCGTCTGATGCC 1230

217 GlnAlaGlyAlaGluLeuThrTyrProSerAlaPheGlySerIleTh 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1231 GACCAAGAAATGCAGATCTCTGTTGTACCGTTTCTACCCGATACACAAA 1280

233 rGlyProAlaHisThrGluValLeuLeuArgAlaArgAlaIleGluThrG 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1281 TGTATTATTCGCGTGTTCGGGTCTCGGCACAGGCACGTCGCATTGAGAACG 1330

250 lnCysTyrValValAlaAlaAlaGlnCysGly.....ArgHisHis 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1331 AATGCTTTGTGTTAATAGCCGCGCAGTGTAGGCAATCTTCCCGTGTGCAC 1380

264 GluLysArgAlaSerTyrGlyHisSerMetValValAspPro..... 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1381 AATATGGATATTCAATATGCTCAGTCGGGAGTATTCACACCTTCGGATT 1430

278 ....TrpGlyThrValValAlaArgCysSerGluGlyProGly..... 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1431 CGTTTTCGACACAGCGAAGCGTGCCGAAGCAACTCCGAATACAGAAA 1480

291 ..LeuCysLeuAlaArgIleAspLeuAsnTyrLeuArgGlnLeuArg 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1481 TGATCTGGTTTCGGATGTAGATCTCGACTTATTTGAACGAACTACACACT 1530

307 HisLeuProVal.....PheGlnHisArgArgProAspLeuTyr 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1531 TAGCGCAGCGTTCCCAACCTGAAGGACAGCGCAAAATGATGATAT 1575
```

seq\_name: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:US-10-102-524-748

```
seq_documentation_block:
; Sequence 748, Application US/10102524
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 748
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-748
```

```
alignment_scores:
  Quality: 129.50      Length: 59
  Ratio: 2.943        Gaps: 1
  Percent Similarity: 74.576  Percent Identity: 42.373

alignment_block:
US-09-357-675C-21 x US-10-102-524-748 ..
Align seg 1/1 to: US-10-102-524-748 from: 1 to: 270

264 GluLysArgAlaSerTyr.....GlyHisSerMetValValAspPr 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
21 GATGACAAAGCCCTCTATGTTGCCCTGGGACACAGCACCGTGGTGAACCC 70

277 oTrpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuA 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

[illegible]

; PRIOR APPLICATION NUMBER: 08/903,802
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/905,881
; PRIOR FILING DATE: August 1, 1997
; PRIOR APPLICATION NUMBER: 60/025,204
; PRIOR FILING DATE: August 1, 1996
; PRIOR APPLICATION NUMBER: 08/903,471
; PRIOR FILING DATE: July 30, 1997
; PRIOR APPLICATION NUMBER: 60/025,478
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/903,556
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/025,217
; PRIOR FILING DATE: August 22, 1996
; PRIOR APPLICATION NUMBER: 08/937,142
; PRIOR FILING DATE: September 23, 1997
; PRIOR APPLICATION NUMBER: 60/026,598
; PRIOR FILING DATE: September 24, 1996
; PRIOR APPLICATION NUMBER: 08/960,746
; PRIOR FILING DATE: October 29, 1997
; PRIOR APPLICATION NUMBER: 60/030,144
; PRIOR FILING DATE: October 30, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/755,524
; PRIOR FILING DATE: November 22, 1996
; PRIOR APPLICATION NUMBER: 60/007,495
; PRIOR FILING DATE: November 22, 1995
; PRIOR APPLICATION NUMBER: 09/021,031
; PRIOR FILING DATE: February 10, 1998
; PRIOR APPLICATION NUMBER: 60/039,325
; PRIOR FILING DATE: February 13, 1997
; PRIOR APPLICATION NUMBER: 09/035,172
; PRIOR FILING DATE: March 4, 1998
; PRIOR APPLICATION NUMBER: 60/040,431
; PRIOR FILING DATE: March 5, 1997
; PRIOR APPLICATION NUMBER: 09/041,894
; PRIOR FILING DATE: March 12, 1998
; PRIOR APPLICATION NUMBER: 60/040,199
; PRIOR FILING DATE: March 14, 1997
; PRIOR APPLICATION NUMBER: 09/050,817
; PRIOR FILING DATE: March 30, 1998
; PRIOR APPLICATION NUMBER: 60/043,792
; PRIOR FILING DATE: April 11, 1997
; PRIOR APPLICATION NUMBER: 09/074,999
; PRIOR FILING DATE: May 8, 1998
; PRIOR APPLICATION NUMBER: 60/048,431
; PRIOR FILING DATE: May 29, 1997
; PRIOR APPLICATION NUMBER: 09/107,592
; PRIOR FILING DATE: June 30, 1998
; PRIOR APPLICATION NUMBER: 60/052,751
; PRIOR FILING DATE: July 1, 1997
; PRIOR APPLICATION NUMBER: 09/094,079
; PRIOR FILING DATE: June 9, 1998
; PRIOR APPLICATION NUMBER: 60/049,975
; PRIOR FILING DATE: June 13, 1997
; NUMBER OF SEQ ID NOS: 35654
; SOFTWARE: PERL Program
; SEQ ID NO 5721
; LENGTH: 252
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc\_feature
; OTHER INFORMATION: Incyte ID No: hu00212121
; NAME/KEY: unsure
; LOCATION: 65, 167
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-210B-5721

alignment\_scores:
Quality: 124.00 Length: 71
Ratio: 2.431 Gaps: 2
Percent Similarity: 71.831 Percent Identity: 38.028

alignment\_block:
US-09-357-675c-21 x US-09-540-210B-5721 ..
Align seg 1/1 to: US-09-540-210B-5721 from: 1 to: 252

252 TyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaSe 268
|||||:|||||: ||| ||| |||||:|||||: |||
6 TATGTGGCCACAGCCTCTCTCTGCC.....CGGGATGACAAAGCCTC 46

268 rTyr.....GlyHisSerMetValValAspProTrrpGlyThrValV 282
|||||:|||||: ||| ||| |||||:|||||: |||
47 CTATGTTGCCTGGGACANAGCACCGTGTGAACCTTGGGGGAGGTTTC 96

282 alAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgilleAspLeu 298
:|||||:|||||: ||| ||| |||||:|||||: |||
97 TAGCCAAAGCTGGCAGACAGAGCAATCGTGTATTCACACATAGACCTG 146

299 AsnTyrLeuArgGlnLeuArgHisLeuProValPheGlnHisArgAr 315
::: ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
147 AAGAAGCTGGCTGAATACGNCAGCAAAATCCCGTTTTTAGACAGAAGCG 196

315 gProAspLeuTyr 319
| |||||
197 ATCAGACCTCTAT 209

seq\_name: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:US-09-540-210B-5426

seq\_documentation\_block:
; Sequence 5426, Application US/09540210B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF URINARY TRACT TISSUE
; FILE REFERENCE: PD-1037 CIP
; CURRENT APPLICATION NUMBER: US/09/540,210B
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/972,899
; PRIOR FILING DATE: November 18, 1997
; PRIOR APPLICATION NUMBER: 08/395,244
; PRIOR FILING DATE: February 27, 1995
; PRIOR APPLICATION NUMBER: 08/722,922
; PRIOR FILING DATE: September 27, 1996
; PRIOR APPLICATION NUMBER: 60/005,526
; PRIOR FILING DATE: September 29, 1995
; PRIOR APPLICATION NUMBER: 08/824,029
; PRIOR FILING DATE: March 25, 1997
; PRIOR APPLICATION NUMBER: 60/014,010
; PRIOR FILING DATE: March 25, 1996
; PRIOR APPLICATION NUMBER: 08/826,847
; PRIOR FILING DATE: April 10, 1997
; PRIOR APPLICATION NUMBER: 60/015,533
; PRIOR FILING DATE: April 10, 1996
; PRIOR APPLICATION NUMBER: 08/903,555
; PRIOR FILING DATE: July 31, 1997
; PRIOR APPLICATION NUMBER: 60/023,308
; PRIOR FILING DATE: July 31, 1996
; PRIOR APPLICATION NUMBER: 08/862,178
; PRIOR FILING DATE: May 22, 1997
; PRIOR APPLICATION NUMBER: 60/018,217
; PRIOR FILING DATE: May 23, 1996
; PRIOR APPLICATION NUMBER: 08/881,589
; PRIOR FILING DATE: June 24, 1997









```
4988 .....CTGCCCCAGTGTGGGA 5003
261 rghHisHisGluLysArgAlaSer.....TyrGlyHisSerMetVal 274
||||| ..... ||||| ||||| |||||
5004 GGCACCTCCATCGCCAGCAGGCTGAGATCAAGACACGGT..... 5047
275 ValAspProTrpGlyThrValAlaAlaArgCysSerGluGlyProGly.. 290
||| ..... ||||| ||||| |||||
5048 .....GCCAGGCTGCAACCTCTGTCCCAAGGGCAGACGAG 5085
291 .....LeuCysLeuAlaArgIleAspLeuAsn 299
||||| ..... ||||| |||||
5086 AAAGCGGCTTGTCTGTCTGCTCGGTTTCTGTGTCC 5119
seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-108
seq_documentation_block:
; Sequence 108, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-108
alignment_scores:
Quality: 92.50 Length: 326
Ratio: 0.638 Gaps: 16
Percent Similarity: 44.479 Percent Identity: 24.233
alignment_block:
US-09-357-675c-21 x US-10-105-299-108 ..
Align seg 1/1 to: US-10-105-299-108 from: 1 to: 1536
6 ThrArgPro...HisArgPheLeuSerLeuLeuCysProGlyLeuAr 21
||| ||||| ||| ||||| |||||
13 ACATCTCCCCATAGCAC.....CCTGCCCTCAT 41
21 gIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMeta 38
||| ||| ||||| ||||| ||||| |||||
42 GGGACCTGCCCTC.....CCTCAGCGTCAGCCATCA 73
38 lalleSerSerSerCysGluLeuProLeuVal..... 49
||| ||||| ||||| ||||| |||||
74 GCCATGCCCTCCAGCGCTCTCTAGCCCTCTTCCCAAGGAGCAGAG 123
50 .....AlaValCysGlnValThrSerThrProAspLys... 60
||||| ||||| ||||| ||||| |||||
124 GTGCCACCGGGGTGCTGTCTCTACCTCTCTGCCCTCTAAAGAT 173
61 .....GlnGlnAsnPheLysThrCysAlaGluLeuValArgGluA 74
||||| ..... ||||| ||||| |||||
174 GGGAGGAGACCGGGTCCATGGTCTGGCTGTGAGTCTCCCTTCAG 223
74 lalaArgLeuGlyAlaCysLeuAlaPhe...LeuProGluAlaPheAsp 89
||| ||||| ||||| ||||| |||||
224 CTGGTCACTAGGCATCACCCCGCTTTGGTTCTTCAGATGCTCTTGGG 273
90 PheIleAlaArgaspProAlaGluThrHisLeuSerGluProLeuGl 106
||||| ..... ||||| ||||| |||||
274 TTCATAGGGCAGTCTCTAGCGGACGGCCCTGACCTCCCG..... 318
```

```
106 yGlyLysLeuLeuGluLutThrThrGlnLeuAlaAlaArgLysGlyLeuT 123
319 .....GCCT 322
123 rPleuSer.....LeuGlyGlyPheHisGluArg 132
||||| ..... ||||| ||||| |||||
323 GGCTTCACTCTCCCTGACGGCTGCCATTTGGTCCACCTTTTCATAGAGAG 372
133 GlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLe 149
373 .....CCTGCTTTGTT 383
149 uAsnSerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysA 166
||||| ..... ||||| |||||
384 ACAAGCTCGGGTCTCCCTCCTCAGCTCGGTTAAGTACCCGAGCCCTCT 433
166 spValGluIleProGly.GlnGlyProMetCysGluSerAsnSerThrMe 182
||| ..... ||||| ||||| |||||
434 CTTAAGATGTCAGGGCCCGAGCGCGGCGACAGCCAGCCCAACCTT 483
182 tProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleG 199
||| ..... ||||| |||||
484 G...GGCCTGGAGATCTCTCCACCCCATCCTAGAGTGTCTGACCTG 530
199 lyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLe 215
||||| ..... ||||| ||||| |||||
531 GGCTTTACGGGGCCCATTCACCGCTCCCAACTTGAGCCTGTGACCT 580
215 uAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAla..... 228
||||| ..... ||||| |||||
581 TGGGACCAAGGGGAGTCTCTCTCTTGTGACTCAGCAGAGGAGCAGTGG 630
229 .....PheGlySerIleThrGlyProAla 236
||||| ..... ||||| |||||
631 CCAGTTACAGAGGGCGGTGGCTGGAGGCTAGCCACCTCCAGCT 680
237 HisTrpGluValLeuLeuArgAlaArgAlaIleGlu.....Th 249
||| ..... ||||| ||||| |||||
681 TTTCTCAGGGTGTCTGTGAGTCCAGGATCTGGAGCAATCTGACCTTC 730
249 rGlnCysTyrValValAlaAlaGlnCysGlyArgHisGluLysA 266
||| ..... ||||| ||||| |||||
731 TCCAAAGGCTCTGTATCAGCTGGCAGTGC.....CACCAA... 767
266 rGlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValVal 282
768 .....TCCCTGGCCATTGGCCCCCAGGGGAGCTGGGCC 800
283 AlaArgCysSerGluGlyProGly 290
||||| ..... ||||| |||||
801 TGCAGGCTGCAGGAGGCGACTGGA 824
seq_name: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:US-10-105-299-1860
seq_documentation_block:
; Sequence 1860, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1860
; LENGTH: 2102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-1860
alignment_scores:
```

•  
•  
•  
•

OM of: US-09-357-675c-21 to: EST:\* out\_format : pfs

Date: Apr 29, 2002 10:01 AM

About: Results were produced by the GenCore software, version 4.5.

Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09357675/runat_29042002_091756_3146/app_query.fasta_1.1208
-DB=EST -QFWT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINWATCH=0.100 -LOOCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09357675 -CGN1_1.6497
-NCPU=6 -ICPU=3 -LONGLOG -DEV TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-357-675c-21

Query length: 327

Database: EST\*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 2545.690000

score\_list:

Sequence	Strd Orig	zscore	EScore	Len	! Document
gb.htc:AK004988	1479.00	2490.04	1.9e-129	2993	! AK004988 Mus musculus adult ma
gb.est1:AL520768	1458.00	2468.05	3.2e-128	890	! AL520768 AL520768 LTI_NFL004.NE
gb.htc:AK002269	1458.00	2463.11	6.0e-128	1374	! AK002269 Mus musculus adult ma
gb.est1:AL529152	1413.00	2391.96	5.5e-124	844	! AL529152 AL529152 LTI_NFL001.NE
gb.est2:BI822844	1357.00	2296.29	1.2e-118	928	! BI822844 603040166F1 NIH_MGC.12
gb.est2:BI822844	1357.00	2296.29	1.2e-118	928	! BI822844 603040166F1 NIH_MGC.12
gb.est2:BI822844	1357.00	2296.29	1.2e-118	928	! BI822844 603040166F1 NIH_MGC.12
gb.est2:BI822844	1357.00	2296.29	1.2e-118	928	! BI822844 603040166F1 NIH_MGC.12
gb.est1:AL529163	1253.00	2119.04	8.7e-109	861	! AL529163 602488424F1 NIH_MGC.18
gb.est1:AL529163	1253.00	2119.04	8.7e-109	861	! AL529163 602488424F1 NIH_MGC.18
gb.est1:AL529163	1253.00	2119.04	8.7e-109	861	! AL529163 602488424F1 NIH_MGC.18
gb.est1:AL529163	1253.00	2119.04	8.7e-109	861	! AL529163 602488424F1 NIH_MGC.18
gb.est2:BG762506	1218.00	2060.75	1.5e-105	764	! BG762506 602733954F1 NIH_MGC.49
gb.est2:BG762506	1218.00	2060.75	1.5e-105	764	! BG762506 602733954F1 NIH_MGC.49
gb.est2:BG762506	1218.00	2060.75	1.5e-105	764	! BG762506 602733954F1 NIH_MGC.49
gb.est2:BG762506	1218.00	2060.75	1.5e-105	764	! BG762506 602733954F1 NIH_MGC.49
gb.est1:AL520767	1201.50	2029.66	8.3e-104	790	! AL520767 602686082F1 NIH_MGC.95
gb.est1:AL520767	1201.50	2029.66	8.3e-104	790	! AL520767 602686082F1 NIH_MGC.95
gb.est1:AL520767	1201.50	2029.66	8.3e-104	790	! AL520767 602686082F1 NIH_MGC.95
gb.est1:AL520767	1201.50	2029.66	8.3e-104	790	! AL520767 602686082F1 NIH_MGC.95
gb.est1:AL523733	1179.50	1992.54	9.7e-102	960	! AL523733 603023373 LTI_NFL004.NE
gb.est1:AL523733	1179.50	1992.54	9.7e-102	960	! AL523733 603023373 LTI_NFL004.NE
gb.est1:AL523733	1179.50	1992.54	9.7e-102	960	! AL523733 603023373 LTI_NFL004.NE
gb.est1:AL523733	1179.50	1992.54	9.7e-102	960	! AL523733 603023373 LTI_NFL004.NE
gb.est2:BI75823	1169.00	1974.81	9.5e-101	946	! BI75823 603030404F1 NIH_MGC.11
gb.est2:BI75823	1169.00	1974.81	9.5e-101	946	! BI75823 603030404F1 NIH_MGC.11
gb.est2:BI75823	1169.00	1974.81	9.5e-101	946	! BI75823 603030404F1 NIH_MGC.11
gb.est2:BI75823	1169.00	1974.81	9.5e-101	946	! BI75823 603030404F1 NIH_MGC.11
gb.est2:BI755493	1135.50	1919.44	1.1e-97	813	! BI755493 603027349F1 NIH_MGC.11
gb.est2:BI755493	1135.50	1919.44	1.1e-97	813	! BI755493 603027349F1 NIH_MGC.11
gb.est2:BI755493	1135.50	1919.44	1.1e-97	813	! BI755493 603027349F1 NIH_MGC.11
gb.est2:BI755493	1135.50	1919.44	1.1e-97	813	! BI755493 603027349F1 NIH_MGC.11
gb.est2:BI711300	1115.00	1887.49	6.9e-96	625	! BI711300 1d96f12.x1 Human insul
gb.est2:BI711300	1115.00	1887.49	6.9e-96	625	! BI711300 1d96f12.x1 Human insul
gb.est2:BI711300	1115.00	1887.49	6.9e-96	625	! BI711300 1d96f12.x1 Human insul
gb.est2:BI711300	1115.00	1887.49	6.9e-96	625	! BI711300 1d96f12.x1 Human insul
gb.est2:BI752623	1109.00	1875.25	3.3e-95	746	! BI752623 603028471F1 NIH_MGC.11
gb.est2:BI752623	1109.00	1875.25	3.3e-95	746	! BI752623 603028471F1 NIH_MGC.11
gb.est2:BI752623	1109.00	1875.25	3.3e-95	746	! BI752623 603028471F1 NIH_MGC.11
gb.est2:BI752623	1109.00	1875.25	3.3e-95	746	! BI752623 603028471F1 NIH_MGC.11
gb.est2:BE895409	1097.50	1854.47	4.8e-94	828	! BE895409 601434778F1 NIH_MGC.72
gb.est2:BE895409	1097.50	1854.47	4.8e-94	828	! BE895409 601434778F1 NIH_MGC.72
gb.est2:BE895409	1097.50	1854.47	4.8e-94	828	! BE895409 601434778F1 NIH_MGC.72
gb.est2:BE895409	1097.50	1854.47	4.8e-94	828	! BE895409 601434778F1 NIH_MGC.72
gb.est2:BI647584	1089.00	1839.92	3.1e-83	832	! BI647584 603278164F1 NIH_MGC.18
gb.est2:BI647584	1089.00	1839.92	3.1e-83	832	! BI647584 603278164F1 NIH_MGC.18
gb.est2:BI647584	1089.00	1839.92	3.1e-83	832	! BI647584 603278164F1 NIH_MGC.18
gb.est2:BI647584	1089.00	1839.92	3.1e-83	832	! BI647584 603278164F1 NIH_MGC.18
gb.est2:BG76560	1087.00	1837.09	4.4e-93	791	! BG76560 603663693F1 NIH_MGC.59
gb.est2:BG76560	1087.00	1837.09	4.4e-93	791	! BG76560 603663693F1 NIH_MGC.59
gb.est2:BG76560	1087.00	1837.09	4.4e-93	791	! BG76560 603663693F1 NIH_MGC.59
gb.est2:BG76560	1087.00	1837.09	4.4e-93	791	! BG76560 603663693F1 NIH_MGC.59
gb.est2:BG532265	1079.50	1824.45	2.2e-92	781	! BG532265 602561355F1 NIH_MGC.61
gb.est2:BG532265	1079.50	1824.45	2.2e-92	781	! BG532265 602561355F1 NIH_MGC.61
gb.est2:BG532265	1079.50	1824.45	2.2e-92	781	! BG532265 602561355F1 NIH_MGC.61
gb.est2:BG532265	1079.50	1824.45	2.2e-92	781	! BG532265 602561355F1 NIH_MGC.61
gb.est2:BG63013	1077.00	1818.98	4.5e-92	869	! BG63013 602828020F1 NCI_CGAP.5
gb.est2:BG63013	1077.00	1818.98	4.5e-92	869	! BG63013 602828020F1 NCI_CGAP.5
gb.est2:BG63013	1077.00	1818.98	4.5e-92	869	! BG63013 602828020F1 NCI_CGAP.5
gb.est2:BG63013	1077.00	1818.98	4.5e-92	869	! BG63013 602828020F1 NCI_CGAP.5
gb.est1:AL562604	1067.50	1803.00	3.5e-91	853	! AL562604 602562604 LTI_NFL003.NE
gb.est1:AL562604	1067.50	1803.00	3.5e-91	853	! AL562604 602562604 LTI_NFL003.NE
gb.est1:AL562604	1067.50	1803.00	3.5e-91	853	! AL562604 602562604 LTI_NFL003.NE
gb.est1:AL562604	1067.50	1803.00	3.5e-91	853	! AL562604 602562604 LTI_NFL003.NE
gb.est2:BF968272	1045.50	1764.89	4.7e-89	900	! BF968272 602266958F1 NIH_MGC.84
gb.est2:BF968272	1045.50	1764.89	4.7e-89	900	! BF968272 602266958F1 NIH_MGC.84
gb.est2:BF968272	1045.50	1764.89	4.7e-89	900	! BF968272 602266958F1 NIH_MGC.84
gb.est2:BF968272	1045.50	1764.89	4.7e-89	900	! BF968272 602266958F1 NIH_MGC.84
gb.est2:BI332807	1041.50	1758.55	1.0e-88	863	! BI332807 602985469F1 NCI_CGAP.11
gb.est2:BI332807	1041.50	1758.55	1.0e-88	863	! BI332807 602985469F1 NCI_CGAP.11
gb.est2:BI332807	1041.50	1758.55	1.0e-88	863	! BI332807 602985469F1 NCI_CGAP.11
gb.est2:BI332807	1041.50	1758.55	1.0e-88	863	! BI332807 602985469F1 NCI_CGAP.11
gb.est2:BG765807	1035.00	1750.91	2.8e-88	638	! BG765807 ux23q06.y1 Soares thym
gb.est2:BG765807	1035.00	1750.91	2.8e-88	638	! BG765807 ux23q06.y1 Soares thym
gb.est2:BG765807	1035.00	1750.91	2.8e-88	638	! BG765807 ux23q06.y1 Soares thym
gb.est2:BG765807	1035.00	1750.91	2.8e-88	638	! BG765807 ux23q06.y1 Soares thym
gb.est2:BG667049	1025.00	1730.04	4.1e-87	893	! BG667049 DRABTCL1 Rat DRG Libra
gb.est2:BG667049	1025.00	1730.04	4.1e-87	893	! BG667049 DRABTCL1 Rat DRG Libra
gb.est2:BG667049	1025.00	1730.04	4.1e-87	893	! BG667049 DRABTCL1 Rat DRG Libra
gb.est2:BG667049	1025.00	1730.04	4.1e-87	893	! BG667049 DRABTCL1 Rat DRG Libra
gb.est2:BF696074	1013.50	1711.32	4.5e-86	827	! BF696074 601852067F1 NIH_MGC.58
gb.est2:BF696074	1013.50	1711.32	4.5e-86	827	! BF696074 601852067F1 NIH_MGC.58
gb.est2:BF696074	1013.50	1711.32	4.5e-86	827	! BF696074 601852067F1 NIH_MGC.58
gb.est2:BF696074	1013.50	1711.32	4.5e-86	827	! BF696074 601852067F1 NIH_MGC.58
gb.est2:BI711712	1006.50	1699.75	2.0e-85	801	! BI711712 602979783F1 NCI_CGAP.11
gb.est2:BI711712	1006.50	1699.75	2.0e-85	801	! BI711712 602979783F1 NCI_CGAP.11
gb.est2:BI711712	1006.50	1699.75	2.0e-85	801	! BI711712 602979783F1 NCI_CGAP.11
gb.est2:BI711712	1006.50	1699.75	2.0e-85	801	! BI711712 602979783F1 NCI_CGAP.11
gb.est1:AL580840	1003.00	1693.38	4.5e-85	579	! AL580840 601850840 LTI_NFL008.TC
gb.est1:AL580840	1003.00	1693.38	4.5e-85	579	! AL580840 601850840 LTI_NFL008.TC
gb.est1:AL580840	1003.00	1693.38	4.5e-85	579	! AL580840 601850840 LTI_NFL008.TC
gb.est1:AL580840	1003.00	1693.38	4.5e-85	579	! AL580840 601850840 LTI_NFL008.TC
gb.est2:BG432389	992.50	1686.49	1.1e-84	900	! BG432389 602495419F1 NIH_MGC.12
gb.est2:BG432389	992.50	1686.49	1.1e-84	900	! BG432389 602495419F1 NIH_MGC.12
gb.est2:BG432389	992.50	1686.49	1.1e-84	900	! BG432389 602495419F1 NIH_MGC.12
gb.est2:BG432389	992.50	1686.49	1.1e-84	900	! BG432389 602495419F1 NIH_MGC.12
gb.est2:BI340308	992.50	1686.49	1.1e-84	900	! BI340308 602495419F1 NIH_MGC.12
gb.est2:BI340308	992.50	1686.49	1.1e-84	900	! BI340308 602495419F1 NIH_MGC.12
gb.est2:BI340308	992.50	1686.49	1.1e-84	900	! BI340308 602495419F1 NIH_MGC.12
gb.est2:BI340308	992.50	1686.49	1.1e-84	900	! BI340308 602495419F1 NIH_MGC.12
gb.est2:BF80562	951.50	1605.73	3.4e-80	821	! BF80562 602103103F1 NCI_CGAP.11
gb.est2:BF80562	951.50	1605.73	3.4e-80	821	! BF80562 602103103F1 NCI_CGAP.11
gb.est2:BF80562	951.50	1605.73	3.4e-80	821	! BF80562 602103103F1 NCI_CGAP.11
gb.est2:BF80562	951.50	1605.73	3.4e-80	821	! BF80562 602103103F1 NCI_CGAP.11
gb.est2:BI392424	939.00	1586.71	3.9e-79	672	! BI392424 601308982F1 NIH_MGC.44
gb.est2:BI392424	939.00	1586.71	3.9e-79	672	! BI392424 601308982F1 NIH_MGC.44
gb.est2:BI392424	939.00	1586.71	3.9e-79	672	! BI392424 601308982F1 NIH_MGC.44
gb.est2:BI392424	939.00	1586.71	3.9e-79	672	! BI392424 601308982F1 NIH_MGC.44
gb.est2:BI322101	931.00	1574.50	1.9e-78	593	! BI322101 maa34b02.y1 NCI_CGAP.11
gb.est2:BI322101	931.00	1574.50	1.9e-78	593	! BI322101 maa34b02.y1 NCI_CGAP.11
gb.est2:BI322101	931.00	1574.50	1.9e-78	593	! BI322101 maa34b02.y1 NCI_CGAP.11
gb.est2:BI322101	931.00	1574.50	1.9e-78	593	! BI322101 maa34b02.y1 NCI_CGAP.11
gb.est2:BM142111	912.00	1542.93	1.1e-76	552	! BM142111 y1 Melton Norm
gb.est2:BM142111	912.00	1542.93	1.1e-76	552	! BM142111 y1 Melton Norm
gb.est2:BM142111	912.00	1542.93	1.1e-76	552	! BM142111 y1 Melton Norm
gb.est2:BM142111	912.00	1542.93	1.1e-76	552	! BM142111 y1 Melton Norm
gb.est2:BF385518	903.00	1523.52	1.3e-75	789	! BF385518 602047063F1 NCI_CGAP.11
gb.est2:BF385518	903.00	1523.52	1.3e-75	789	! BF385518 602047063F1 NCI_CGAP.11
gb.est2:BF385518	903.00	1523.52	1.3e-75	789	! BF385518 602047063F1 NCI_C

Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5'-GAGAGAGAGCGCGCCACACGACGAGCTTTTCTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. Second  
 strand cDNA was prepared with the primer adapter of sequence [5'  
 GAGAGAGAGAGGATCCACAGGCTCAATTAATTAACCCGCCCC 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOLR.

## FEATURES

Location/Qualifiers

1..2993  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGI:1895066"  
 /db\_xref="taxon:10090"  
 /clone="I300013F05"  
 /sex="male"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 169..1041  
 /gene="Nt1"  
 169..1041  
 /gene="Nt1"  
 /note="data source:MGI, source key:MGI:1350916,  
 evidence:ISS  
 nitrilase 1  
 putative"  
 /codon\_start=1  
 /protein\_id="BAB23723.1"  
 /db\_xref="GI:12836591"  
 /translation="MSSSTSWELPVAVCQVTSIPNKOENFKTCAELVQEAARLGACL  
 AFLPEAFDIARNFAETILLSEPLNGDLGOY SOLARECGIWLISLGFHERGQWQON  
 QKIYNCHVLLSKGSVAVSYRTHLCDVEIPGQPMRESNYTKPGGLEPPVKTPAGK  
 VGLAICVDMPRELSLQAQAEILTPSAGSVTPAHWEVLLRARAIRESQCYVIA  
 AAQGRHETRAISYGHSMVDPWGTVMVAVRCSEGPLCLARIIDLHFLQOMRHLPLVFOH  
 RRPYLGSLSGLPLS"  
 2973..2978  
 /note="putative"  
 2993  
 /note="putative"  
 785 a 739 c 747 g 722 t

gene

CDS

polyA\_signal

polyA\_site

BASE COUNT 785 a 739 c 747 g 722 t  
 ORIGIN

alignment\_scores:

Quality: 1479.00 Length: 326  
 Ratio: 4.833 Gaps: 2  
 Percent Similarity: 93.865 Percent Identity: 84.356

alignment\_block:

US-09-357-675c-21 x AK004988 ..

Align seg 1/1 to: AK004988 from: 1 to: 2993

2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCyspr 18  
 |||||  
 73 CTGGCTTATCATCAGGCTCTCTACCAA.....CTCCGTGTAC 113  
 18 oGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgproA 35  
 |||||  
 114 CGGATACCGATTACTTCGAACCCCACTTGTGTACTACGAGCCAGGCCA 163  
 35 rgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAlaVal 51  
 |||||

164 GAACCATGTCC...TCATCAACTTCTCGGAGGTGCCCTGGTGGCTGTG 210  
 52 CysGlnValThrSerThrProAspLysGlnGlnAsnPhelYsThrCysal 68  
 |||||  
 211 TGGCAGGTAACTCAACCAACAAGCAAGAGACTTTAAACACATGTGC 260  
 68 aGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuP 85  
 |||||  
 261 TGAGTTGGTTCAGAGGCTGCCAGACTGGTGTGCTTGCCTGGCTTCTGC 310  
 85 roGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101  
 |||||  
 311 CTGAGGCATTTGACTTTATTCAGCAAAACCTGCCGAGACATTTACTCCTG 360  
 102 SerGluProLeuGlyLysLeuLeuGluGlyThrGlnLeuAlaAr 118  
 |||||  
 361 TCGGAACCACTGAATGGGATCTTTTGGGCCAATATAGCCAGCTGCCAG 410  
 118 gGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnA 135  
 |||||  
 411 GGAATGTGCAATCTGGCTCTCTTGGCGGTTTCCACGAGCGTGCCCAAG 460  
 135 spTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsnSer 151  
 |||||  
 461 ACTGGGAGCAGAATCAGAAAATCTACAATTTGTCATGTGCTTTTGAACAGC 510  
 152 LysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspValGl 168  
 |||||  
 511 AAGGGATCAGTAGTGCCAGTTACAGGAAGACACATCTGTGGCATGTAGA 560  
 168 uIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProGlyP 185  
 |||||  
 561 GATCCAGCTCAGGGGCCGATGAGAGAAGCAACTATATCAAGCCTGGAG 610  
 185 roSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201  
 |||||  
 611 GCACCTCTTGAGCCACCTGTCAAGACACCGCTGGCAAGTTGTCTAGCA 660  
 202 ValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAl 218  
 |||||  
 661 ACTGTGTATGACATGGGTCTCCCTGAACCTTCTTTGAAATGGCTCAAGC 710  
 218 aglyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyP 235  
 |||||  
 711 TGGGGCAGAAATACTTACTTATCTTTCAGCCTTTGGATCTGTACAGGTC 760  
 235 roAlaHisTrpGluValLeuLeuArgAlaAlaIleGluThrGlnCys 251  
 |||||  
 761 CAGCCCACTGGGAGTGTGCTGCGGGCCCGCCATTTGAATCTCAGTGC 810  
 252 TyrValValAlaAlaGlnCysGlyArgHisGluLysArgAlaSe 268  
 |||||  
 811 TATGTAATAGCAGCGCGATGTGGACGCCACCATGAACAAGAGCAAG 860  
 268 rTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgC 285  
 |||||  
 861 TTATGGCCATAGCATGGTGTGTTACCCCGTGGGCACAGTGGTGGCCCGCT 910  
 285 ysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 301  
 |||||  
 911 GTCGAGGAGGACGAGGCTCTGCTTGTGCTGAATTTGATCTCCACTTTCTA 960  
 302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLe 318  
 |||||  
 961 CAACAGATGCGCCACACCTGCTGTGTTCCAGACCCGACGACCTGACCT 1010  
 318 utyGlyAsnLeuGlyHisProLeuSer 327  
 |||||  
 1011 CTATGGCAGTCTGGGTCTATCCACTCTCT 1038  
 seq\_name: gb\_est1:AL520768  
 seq\_documentation\_block:

LOCUS	AL520768	890 bp	mRNA	linear	EST 13-FEB-2001			
DEFINITION	AL520768	LTI_NFL004_NBC2	Homo sapiens	cdna clone	CSODB0021D05.5			
ACCESSION	AL520768	prime, mRNA sequence.						
VERSION	AL520768							
KEYWORDS	AL520768.1	GI:12784261						
SOURCE	EST.							
ORGANISM	human.							
	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE	1	(bases 1 to 890)						
AUTHORS	Li, W. B., Gruber, C., Tesse, J. and Polayes, D.							
TITLE	Full-length cDNA libraries and normalization							
JOURNAL	Unpublished (2001)							
COMMENT	Contact: Genoscope							
	Genoscope - Centre National de Sequencage							
	BP 191 91006 EVRY cedex - France							
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.							
FEATURES	Location/Qualifiers							
source	1..890							

```

location/quad111115
1. .830
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CSDB002YD05"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
191 a 251 c 241 g 206 t 1 others
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 1458.00
  Ratio: 5.302
  Percent Similarity: 99.278
  Length: 277
  Gaps: 0
  Percent Identity: 99.278

alignment_block:
  US-09-357-675C-21 x AL520768

```

```

Align seg 1/1 to: AL520768 from: 1 to: 890

1 MetLeuGlyPheIleThrArgProHisIArGheLeuSerLeuLeuCy 17
|||||
60 ATGCTGGCGTCATCACAGCGCTCCTTCACAGATTCTGTGCCCTTCTGTG 109
|||||
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34
|||||
110 TCCCTGGAGCTCGGATACCTCAACTCTCAGTACTTTGTGTACGCCAGGC 159
|||||
34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50
|||||
160 CCAGAGCCATGGCTATCTCTCTCTCTCGAAGTCGCCCTGGTGGCT 209
|||||
51 ValCysGlnValThrSerThrProAspGlySglnGlnAsnPhelysthrCy 67
|||||
210 GTGTGCCAGGTAAACATCGAGCCAGACAAGCAACAGAACTTTAAACATG 259
|||||
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
260 TGCTGAGCTGGTTTCGAGAGGCTGCCAGACTGGGTGCCTGCCTGGCTTCC 309
|||||
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||

```

310	TCGCTGAGGCAATTGACTTCATTTGCACGGGACCTGCAGAGACGGCTACAC	355
101	LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAla	117
360	CTGCTCTGAACCACTGGGTGGAAACTTTTGAAGAATACACCCAGCTTGC	409
117	aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG	134
410	CAGGGAATGTGGACTCTGGCTGCTCTTGGGTGGTGTTCATGAGCGTGCC	459
134	lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn	150
460	AGAAGTGGGAGCAGACTCAGAAATCTACAAATTGTACGCTGCTGCTGAAC	509
151	SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa	167
510	AGCAAGGGCAGCTAGTGGCCACTTACAGGAAGACACATCTGTGTGACGT	559
167	lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG	184
560	AGAGATTCACGGCAGGGGCGCTATGTGTGAAAGCAACTCTACCATGCCTG	609
184	lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu	200
610	GGCCCCAGTCTTGAGTCACCTGTACAGCACACAGCAGGCAAGATTGGTCTA	659
201	AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG	217
660	GCTGTCTGCTATCACATCGCGTCCCTGAACTCTCTGTGGCATTTGGCTCA	709
217	nAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThrG	234
710	AGCTGGAGCAGAGATACCTTACCTATCCTTCAGCTTTTGGATCCATTACAG	759
234	lyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGln	250
760	GGCCAGCCCACTGGGAGGTGTGTGGCGGCCCGCTGTATCGAAACCCAG	809
251	CysTyrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAl	267
810	TGCTATGTAGTGGCAGCAGCAGCAGTGTGGAGCCCACTAGAAAGAGAGC	859
267	asertYrGlyHisSerMetValValAspPro	277
860	AAGTTATGGCCACAGCATGTTGGKAGACCCC	890
seq_name: gb_htc:AK002269		
seq_documentation_block:		
LOCUS	AK002269	1374 bp mRNA linear
DEFINITION	Mus musculus adult male kidney cDNA, RIKEN full-length library, clone:0610007B07:nitrilase 1, full insert.	
ACCESSION	AK002269	
VERSION	AK002269.1	GI:12832130
KEYWORDS	HTC; CAP trapper.	
SOURCE	Mus musculus (strain:C57BL/6J) adult male kidney cDNA clone.lib:RIKEN full-length enriched mouse cDNA lib clone:0610007B07.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae	
AUTHORS	1 (sites)	
TITLE	High-efficiency full-length cDNA cloning	
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE	99279253	
PUBMED	10349636	
REFERENCE	2 (sites)	
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., S. Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashi, K.	
TITLE	Normalization and subtraction of cap-trapper-select prepare full-length cDNA libraries for rapid discovery	
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	





235 ProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluThrGlnCys 251  
|||||  
807 CCAGCCCACTGGGAGGTGCTGTCGGGGCCGCCGCAATGAATCTCAGTG 856  
  
251 sTyTrValValAlaAlaAlaGlnCysGlyArgHisHisGluLysArgAlaS 268  
|||||  
857 CTAATGTAATAGCAGCAGCCAGCTGTGGAGCCGCCACCATGAACAAGACAA 906  
  
268 exTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArg 284  
|||||  
907 GTATGGCCATAGCTGGTGGTGACCCGTGGGCACAGCTGGTGGCCCG 956  
  
285 CysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLe 301  
957 TGTCTCCGAGGGACCGAGGCTGTGCTTGTGCGAATGTGATCTCCACTTTCT 1006  
  
301 uArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspL 318  
|||||  
1007 ACAACAGATGCGCCCAACACTGCCTGTGTTCAGCACCGCCAGACCTTGACC 1056  
  
318 euTyrGlyAsnLeuGlyHisProLeuSer 327  
|||||  
1057 TCTATGGCACTGGGTCTCATCCACTCTCT 1085

seq\_name: gb\_est1:AL529152

seq\_documentation\_block:  
LOCUS AL529152 844 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL529152 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DD004YM06 5  
prime, mRNA sequence.

ACCESSION AL529152  
VERSION AL529152.1 GI:12792645  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 844)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Source

1..844

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="LTI\_NFL001\_NBC4"

/sex="male"

/tissue\_type="neuroblastoma cells"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-Oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with NotI and

cloned into the NotI and EcoRV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed

by Life Technologies. Contact : Feng Liang Life

Technologies, a division of Invitrogen 9800 Medical Center

Drive Rockville, Maryland 20850, USA Fax : (1) 301 610

8371 Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

182 a 238 c 235 g 189 t

BASE COUNT

ORIGIN

alignment\_scores:

Quality: 1413.00

Ratio: 5.373

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 263

Gaps: 0

Percent Identity: 100.000

alignment\_block:

US-09-357-675c-21 x AL529152

Align seg 1/1 to: AL529152 from: 1 to: 844

65 LysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLe 81  
|||||  
1 AAACATGTGCTGAGCTGGTTCGAGAGGCTGCAGACTGGTGGCTGCCT 50  
  
81 uAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluT 98  
|||||  
51 GCGTTTCCTGCCTGAGGCATTTGACTTCATTGCACGGGACCCCTGCAGAG 100  
  
98 hrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThr 114  
|||||  
101 CGGTACACCTGCTGAACCACTGGGTGGGAAACTTTTGAAGAATAACACC 150  
  
115 GlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisG1 131  
|||||  
151 CAGCTTGCCAGGGAATGTGGACTCTGGCTGCTCTTGGTGGTTTCCATGA 200  
  
131 uArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValL 148  
|||||  
201 GCGTGGCCCAAGACTGGGAGCAGACTCAGAAATCTACAATTGTACGTGC 250  
  
148 euLeuAsnSerLysGlyAlaValAlaValAlaThrTyrArgLysThrHisLeu 164  
|||||  
251 TCGTGAACAGCAAGAGGGCAGTAGTGGCCACTTACAGGGAAGACACATCTG 300  
  
165 CysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerTh 181  
|||||  
301 TGTGACCTAGAGATTCCAGGGCAGGGGCTATGTGTGAAAGCAACTCTAC 350  
  
181 rMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysI 198  
|||||  
351 CATGCCCTGGCCCGAGTCTTGAAGTCACTCTCAGCACACAGCAGGCAAGA 400  
  
198 leGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuAla 214  
|||||  
401 TTGGTCTAGCTGCTCTGCTATGACATCGGTTCCCTGAACCTCTCTGGCA 450  
  
215 LeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySe 231  
|||||  
451 TTGGCTCAAGCTGGAGCAGAGATACTTACCTATCTTCAGCTTTTGGATC 500  
  
231 rIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleG 248  
|||||  
501 CATTACAGGCCAGCCCACTGGGAGGTGTGCTGGGGCCCGCTGCTATCG 550  
  
248 luThrGlnCysTyrValValAlaAlaGlnCysGlyArgHisHisGlu 264  
|||||  
551 AAACCCAGTGTCTATGTAGTGGCAGCAGCAGTGTGGAGCCCAACATGAG 600  
  
265 LysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrVa 281  
|||||  
601 AAGAGAGCAAGTATTGGCCACAGCATGGTGGTAGACCCCTGGGGAACAGT 650  
  
281 lValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspL 298  
|||||  
651 GTGGCCCGCTGCTGTGAGGGGCCAGGCCCTCTGCTTGGCCCGAATAGACC 700  
  
298 euAsnTyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArg 314  
|||||  
701 TCNACTATCTGCGACAGTTCGCCGACACCTGCCTGTGTTCAGCACCGC 750  
  
315 ArgProAspLeuTyrGlyAsnLeuGlyHisProLeuSer 327  
|||||  
751 AGGCCTGACCTCTATGGCAATCTGGGTCTACCCCACTGCTCT 789

seq\_name: gb\_est2:BI822844

seq\_documentation\_block:

LOCUS BI822844

928 bp mRNA linear

EST 04-OCT-2001

```
DEFINITION 603040166F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5180897 5',
            mRNA sequence.
ACCESSION  BI822844
VERSION    BI822844.1 GI:15934394
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 928)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LAM11451 row: d column: 18
            High quality sequence stop: 822.
FEATURES   Location/Qualifiers
            1..928
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:5180897"
             /lab_host="DH10B"
             /note="Organ: pooled brain, lung, testis; Vector:
             pCMV-SPO76; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
             source anonymous pool of 6 male brains, age range 23-27; 1
             male lung, age 27; and 1 male testis, age 69. Library is
             oligo-dr primed and directionally cloned (EcoRV site is
             destroyed upon cloning). Average insert size 1.8 kb,
             insert size range 1-3 kb. Library is normalized and
             enriched for full-length clones and was constructed by C.
             Gruber (Invitrogen). Research Genetics tracking code
             021. Note: this is a NIH_MGC Library."
BASE COUNT 198 a 256 c 259 g 214 t 1 others
ORIGIN

alignment_scores:
  Quality: 1357.50      Length: 295
  Ratio: 4.848          Gaps: 4
  Percent Similarity: 94.915 Percent Identity: 93.559

alignment_block:
US-09-357-675C-21 x BI822844 ..

Align seg 1/1 to: BI822844 from: 1 to: 928

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17
|||||
39 ATGCTGGGCTTCATCACCAGGCTCTCTCACAGATTCTCTCCCTCTCTG 88
|||||
17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgp 34
|||||
89 TCCTGGACTCCGGATACCCCAACTCTCAGTACTTTGTGTCAGCCAGGC 138
|||||
34 roArgAlaMetAlaIleSerSerSerCysGlnLeuProLeuValAla 50
|||||
139 CCAGACCAATGCTATCTCTCTCTCTGCACTGCCCTGGTGGCT 188
|||||
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelYsThrCy 67
|||||
189 GTGTGCCAGTAACTCGACGCGCAGACCAAGCAACAACTTTAAACATG 238
|||||
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
```

```
239 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGGCTGCCTGGCTTTC 288
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100
|||||
289 TGCCCTGAGGCATTTGACTTCATTCACGGGACCTTCAGAGACGCTACAC 338
|||||
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
|||||
339 CTGTCTGAACCACTGGTGGGAAACTTTTGAAGAATAACACCCAGCTTGC 388
|||||
117 aArgGlyCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
|||||
389 CAGGAAATGTGACTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 438
|||||
134 InAspTrpGluGlnThrGlnLysIleTyrAspCysHisValLeuLeuAsn 150
|||||
439 AAGACTGGAGCAGACTCAGAAAATCTACAATGTACAGTGTCTGCTGAAC 488
|||||
151 SerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspVa 167
|||||
489 AGCAAGGGCAGTAGTGCCACTTACAGGAAGACACACTCTGTGTGACGT 538
|||||
167 IcluleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
|||||
539 AGAGATTCAGGCAG. GGGCCTATGTGTGAAGCAACTCTACCATGCCTG 587
|||||
184 IyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
588 GGCCAGAGCTTGTAGTCACCTGTACGCCACCCAGCAGGCAAGATTGGTCTA 637
|||||
201 AlavalCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGl 217
|||||
638 GCTGCTCTGATGACATGGGTTCCTGAACTCTCTGGCATTGGCTCA 687
|||||
217 nAlaGlyAlaGluIleLeuThr. TyrProSerAlaPheGlySer...Ile 232
|||||
688 AGCTGGAGCAGAGATACCTTATCTTTCAGCTTTTGGATCCCATATA 737
|||||
233 ThrGlyProAlaHisTrpGluValLeuLeuArgAlaArgAlaIleGluTh 249
|||||
738 CAGGCCCCAGCCACTGGGAGGTGTGTGGGGNCCGTGCTATTCGAAC 787
|||||
249 rGlnCysTyrVal. ValAlaAlaGlnCysGly. ArgHisHisGluLyl 265
|||||
788 CCAGTGTCTATGTTAGTGGCGCCAGCACAGTGTGGAACGCCACCATGAGAA 837
|||||
265 sArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValv 282
|||||
838 GAGAGC. AGTTATGGGCACAGCATGTGTAAACCTTGGGGACAGGGGG 886
|||||
282 alAlaArgCysSerGluGlyProGlyLeu 291
..:
887 CCCGTGCTCTCAGGGCAGGCCCTGGCTTG 915

seq_name: gb_est2:BI769604
seq_documentation_block:
LOCUS      BI769604
DEFINITION 603040166F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:520469 5',
            mRNA sequence.
ACCESSION  BI769604
VERSION    BI769604.1 GI:15761182
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 846)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
```

Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM1512 row: j column: 22  
High quality sequence stop: 838.  
Location/Qualifiers  
1. .846  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5204469"  
/clone\_lib="NIH\_MGC\_122"  
/lab\_host="DH10B"  
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH\_MGC Library."  
BASE COUNT 179 a 240 c 226 g 201 t  
ORIGIN

alignment\_scores:  
Quality: 1308.00 Length: 264  
Ratio: 4.992 Gaps: 2  
Percent Similarity: 99.242 Percent Identity: 99.242  
alignment\_block:  
US-09-357-675c-21 x BI769604 ..  
Align seg 1/1 to: BI769604 from: 1 to: 846  
1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17  
|||||  
59 ATGCTGGGTTTCATCACAGGCTCTCTACAGATTCCTGTCCTTCTGTG 108  
17 sProGlyLeuArgIleProGlnLeuSerValLeuCyAlaGlnProArgP 34  
|||||  
109 TCCTGGACTCCGATACCTCACTCTCAGTACTTTGTGCTCAGCCAGGC 158  
34 roArgIleMetAlaIleSerSerSerCysGluLeuProLeuValAla 50  
|||||  
159 CCAGAGCCATGGCTATCTCTCTCTCTCGGAAGTGCCTGGTGGCT 208  
51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelysThrCy 67  
|||||  
209 GTGTGCAGGTATACATCGAGCCAGACAGAACACAACTTTAAACATG 258  
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
|||||  
259 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTGCTTCC 308  
84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
309 TGCTGAGGCATTTGACTTCATTGACGGGACCCCTGCAGAGAGCTTACAC 358  
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluTyrThrGlnLeuAl 117  
359 CTTGCTGAACCACTGGTGGGAAACATTTTGGAGAATAACCCAGCTTC 408  
117 aArgGlyCysGlyLeuThrLeuSerLeuGlyGlyPheHisGluArgGlyG 134  
409 CAGGGAATGTGGACTCTGGCTGTCTTGGTGGTTCATGAGCGTGGCC 458  
134 InAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150

|||||  
459 AAGACTGGGAGCAGACTCAGAAAAATCTACAATGTGCACGTGCTGAAC 508  
151 SerLysGlyAlaValAlaIleThrTyrArgLysThrHisLeuCysAspVa 167  
|||||  
509 ACAGAAAGGGCAGTAGTGCCACTTACAGAGACACATCTGTGTGACGT 558  
167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
|||||  
559 AGAGATTCCAGGCGAGGGCCCTATGTGTGAAAGCAACTCTACCATGCCTG 608  
184 lYProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
|||||  
609 GGCCCAAGTCTTGAGTCACCTGTGCAGCACAGGCAAGGATTTGCTCTA 658  
201 AlaVal.CysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
|||||  
659 GCTGTGTCATGTATGACATCGGTTCCCTGAATCTCTCTGGCATTTGGCTC 708  
217 InAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233  
|||||  
709 AAGCTGGAGCAGAGATACCTTACCTATCTTCAGCTTTTGGATCCATTACA 758  
234 GlyProAlaHisTrpGluValLeuLeuArgAlaAlaIleGluThrGlu 250  
|||||  
759 GGCCCCAGC.CACTGGGAGGTGTGTGTCGGGC.CGTGCTATCGAAACCA 806  
250 nCysTyrValValAlaAlaAlaGlnCysGlyArgHis 262  
|||||  
807 GTGCTATGTAGTGGCAGCCAGCACAGTGTGGAGCCAC 844  
seq\_name: gb\_est2:BG436916  
seq\_documentation\_block:  
LOCUS BG436916 793 bp mRNA linear EST 14-MAR-2001  
DEFINITION 602488424F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4620892 5',  
mRNA sequence.  
ACCESSION BG436916  
VERSION BG436916.1 GI:13343422  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 793)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP/Gazdar  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1382 row: o column: 05  
High quality sequence stop: 767.  
Location/Qualifiers  
1. .793  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4620892"  
/clone\_lib="NIH\_MGC\_18"  
/tissue\_type="large cell carcinoma"  
/note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and

FEATURES  
source

Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

BASE COUNT 181 a 210 c 230 g 172 t

ORIGIN

alignment\_scores:

Quality: 1298.00 Length: 267  
Ratio: 4.992 Caps: 5  
Percent Similarity: 97.378 Percent Identity: 95.880

alignment\_block:

US-09-357-675C-21 x BG436916 ..

Align seg 1/1 to: BG436916 from: 1 to: 793

46 LeuProLeuValAlaValCysGlnValThrSerThrProAspLysGlnCl 62  
|||||  
3 CTGCCCTGGTGGCTGTGTGCCAGGTAAACATGACGCCAGACAACCA 52  
|||||  
62 nAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGly 79  
|||||  
53 GAACCTTAAACATGTCTGAGCTGGTTCAGAGGCTGCCACACTGGGTG 102  
|||||  
79 laCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspPro 95  
|||||  
103 CCTGCCCTGGCTTCCCTGCCAGGCAATTTGACTTCATTGCACGGGACCT 152  
|||||  
96 AlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeuGluCl 112  
|||||  
153 GCAGAGACGCTACACGTCTGAACACACTGGGTGGGAAACTTTTGGAGA 202  
|||||  
112 uTyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyP 129  
|||||  
203 ATACACCCAGCTGCCAGGAATGTGGACTCTGGCTGTCTTGGGTGGTT 252  
|||||  
129 heHisGluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrrAsnCys 145  
|||||  
253 TCATCAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAAAATCTACAATTGT 302  
|||||  
146 HisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrrArgLysTh 162  
|||||  
303 CAGCTGCTGCTGAACAGCAAGAGGGGAGTAGTGGCCACTTACAGGAAGAC 352  
|||||  
162 rHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerA 179  
|||||  
353 ACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCTATGTGTGAAGCA 402  
|||||  
179 snSerThrMetProGlyProSerLeuGluSerProValSerThrProAla 195  
|||||  
403 ACTCTACCATGCTGGGCCAGTCTTGTAGTCACCTGTCAGCACACACGCA 452  
|||||  
196 GlyLysIleGlyLeuAlaValCysTyrrAspMetArgPheProGluLeuSe 212  
|||||  
453 GGCAAGATTGGTCTAGCTGTCTGTATGACATGCGGTTCCTCGAACTCTC 502  
|||||  
212 rLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrrPro.SerAla 228  
|||||  
503 TCTGGCATTTGGCTCAAGCTGGAGCAGAGATACCTATCTTACAGCT 552  
|||||  
229 PheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuArgAla.A 245  
|||||  
553 TTTGGATTCATTACAGGCCAGCCCACTGGGAGGTGTTGCTCGGGGCCCC 602  
|||||  
245 rAlaIleGluThrGlnCysTyrrValValAlaAlaGlnCysGlyArg 261  
|||||  
603 GTGCTATCGAAACCCAGTCTATGTAGTGGCAGCAGACAGGTGTGACGC 652  
|||||  
262 HisHisGluLysArgAlaSerTyrrGlyHisSerMetValValAsp.ProT 278  
|||||  
653 CACCATGAGAGAGACAGTAAGTTATGGCCACACGATGGTGGTAGACCCCT 702  
|||||  
278 rpGly.ThrValValAlaArgCysSer.GluGlyProGlyLeuCysLeuA 294  
|||||

|||||  
703 GGGAAACAGTGGTGGCGCTGTCTTTGAGGGCCAGGCTCTGCTTGC 752  
|||||

294 laArgIleAspLeuAsnTyrrLeuArgGlnLeuArgArgHis 307  
|||||

753 CCCGAATAGACCTCAAGTATCTGCGACAGTTGCGCCGAAC 793  
|||||

seq\_name: gb\_est1:AL559163

seq\_documentation\_block:

LOCUS AL559163 861 bp mRNA linear EST 16-FEB-2001  
DEFINITION AL559163 LTI\_NFL008\_TC2 Homo sapiens cDNA clone CS0DJ014YA17 5  
prime, mRNA sequence.

ACCESSION AL559163 GI:12904391

VERSION AL559163

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 861)

AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..861

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DJ014YA17"

/clone\_lib="LTI\_NFL008\_TC2"

/sex="male"

/tissue\_type="T cells from T cell leukemia"

/note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 176 a 229 c 225 g 201 t 30 others

ORIGIN

alignment\_scores:

Quality: 1253.00 Length: 280

Ratio: 4.838 Caps: 4

Percent Similarity: 92.500 Percent Identity: 89.286

alignment\_block:

US-09-357-675C-21 x AL559163 ..

Align seg 1/1 to: AL559163 from: 1 to: 861

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuCy 17

|||||

28 ATGCTGGGGCTCATCACCAGGCTCTCACAGATTCCTGTCCCTTCTGTG 77

|||||

17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34

|||||

78 TCCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTCTCAGCCAGGC 127

|||||

34 roArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50

|||||

128 CCAGAGCCATGGCTATCTCCTTCTCCTCGGAACCTGCCCTGGTGGCT 177

|||||

51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThrCy 67

|||||

```
|||||
178 GTGTGCCAGTAATCATGCGCCAGAGTAAGCAAYAGAACTTTAAACATG 227
|||||
67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84
|||||
228 TGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTTCC 277
|||||
84 euProGluAlaPheAspPheAlaAlaArgAspProAlaGluThrLeuHis 100
|||||
278 TGCTGAGGCATTTGACTTTCATTTGATGGAGCCCTGCAGAGAGCTACAY 327
|||||
101 LeuSerGluPro..LeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuA 117
|||||
328 CTGTCTGAACCTACTTGGGTGGGAACATTTTGGAGAAATACAYCCAGCTTK 377
|||||
117 laArgGlu..CysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgG1 133
|||||
378 CCAGGGATATGTGGACTCTKGTCTTCTTGGGTGGTTTCCATGAGCGTGG 427
|||||
133 yGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuA 150
|||||
428 CYAWGACTGGGAGCAGACTTACAAATTTGACATTTGTCACGTGCTGCTGA 477
|||||
150 snSerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAsp 166
|||||
478 ACAGCAAGGGGAGTAGTGGCCACTTACAGGAAGACACATCTCTGTGTGAC 527
|||||
167 ValGluLeuProGlyGlnGlyProMetCysGluSerAsnSerThrMetPr 183
|||||
528 GTAGAGATTCCAGGGCAGGGCCCTATGTGTGAAGCAACTCTACCATGCC 577
|||||
183 oGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyL 200
|||||
578 TGGCCCAAGTCTTTCAGTCACTGTGCAGCACACAGCAGCAAGATTGGTC 627
|||||
200 euAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAla 216
|||||
628 TAGCTGTCTGTATGACATGGGTTCCTGCACCTCTCTCTCRCATTGGCT 677
|||||
217 GlnAlaGlyAlaGluLeuLeuTyrProSerAlaPheGlySerIle..T 233
|||||
678 CMAGCTGGCGCAGAGAWACTTACTTACTTACTTACTTACTTACTTACTT 727
|||||
233 hrGlyProAlaHisTrrp..GluValLeuLeuArgAlaAlaLeuGluTh 249
|||||
728 CARGCCAGCCCACTTGGGAGGTGTTCTGCGGGCCCGTCTCWCMAAM 777
|||||
249 rGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisGluLysA 266
|||||
778 CCAGTGTCTTMTGTAGTGCWCAGCAGCAGTGTGGACGCCACMATGAGA 827
|||||
266 rGlaSerTyrGlyHisSerMetValValAsp 276
|||||
828 GAGCAAGTATATGGCCACAGCATGGTGGTAGAC 859
```

seq\_name: gb\_est2:BG762506

```
seq_documentation_block:
LOCUS      BG762506                      764 bp    mRNA    linear    EST 15-MAY-2001
DEFINITION 502733954F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4859292 5',
            mRNA sequence.
ACCESSION  BG762506
VERSION    BG762506.1  GI:14073159
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 764)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
```

Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1714 row: 1 column: 13  
High quality sequence stop: 749.  
Location/Qualifiers  
1..764  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4859292"  
/clone\_lib="NIH\_MGC\_49"  
/tissue\_type="melanotic melanoma, high MDR (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOT87; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 162 a 218 c 201 g 183 t  
ORIGIN

alignment\_scores:

Quality: 1218.00 Length: 240  
Ratio: 5.183 Gaps: 2  
Percent Similarity: 97.917 Percent Identity: 97.917

alignment\_block:

US-09-357-675C-21 x BG762506 ..

Align seg 1/1 to: BG762506 from: 1 to: 764

```
1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17
|||||
49 ATGCTGGGCTTCATCAGCAGGCTCTCACAGATTCTCTGCTCTCTGTG 98
|||||
17 sProGlyLeuArgIleProGln..LeuSerValLeuCysAlaGlnProArg 33
|||||
99 TCCTGGACTCCGGATACCTCTAACTCTCAGTACTTGTGCTCAGCCACAG 148
|||||
34 ProArgAlaMetAlaIleSerSerSerCysGluLeuProLeuValAla 50
|||||
149 CCCAGAGCCATGGCTATCTCTCTCTCTCCGCACTGCCCTCTGGTGGC 198
|||||
50 aValCysGlnValThrSerThrProAspLysGlnGlnAsnPheLysThr 67
|||||
199 TGTGTGCCAGGTAACATCGAGCCAGACAGCAACAGAACTTTAAACAT 248
|||||
67 yAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPhe 83
|||||
249 GTGCTGAGCTGGTTCGAGAGGCTGCCAGACTGGGTGCTGCTGCTGCTTC 298
|||||
84 LeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHi 100
|||||
299 CTGCTGAGGCATTTGACTTTCATTTGACGGGACCCCTGCAGAGAGCTACA 348
|||||
100 sLeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuA 117
|||||
349 CCTGCTGAACACTGGGTGGGAACTTTGGGAAGAAATACACCCAGCTTG 398
|||||
117 laArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGly 133
|||||
399 CCAGGGAATGTGGACTGTGCTGCTCTGCTTGGGTGGTTTCCATGAGCGTGGC 448
```

134 GlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAs 150  
 |||||  
 449 CAGACTGGGAGCAGACTCAGAAATCTACAATTGTACACGTGCTGAA 498  
 150 nSerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspV 167  
 |||||  
 499 CAGCAAGGGGCGAGTAGTGCCACTTACAGGAAGACACATCTGTGACG 548  
 167 aGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetPro 183  
 |||||  
 549 TAGAGATTCCAGCCAGGGGCCATGTGTGAAGCAACTCTACCATGCCT 598  
 184 GlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLe 200  
 |||||  
 599 GGGCCAGCTTTCAGTCACTGTGCAGACACAGCAGGCAAGATTGGTCT 648  
 200 uAlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
 |||||  
 649 AGCTGTCTGTATGACATGGGGTTCCTGAACCTCTCTGGCATTTGGCTC 698  
 217 lnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleThr 233  
 |||||  
 699 AAGCTGGAGCAGAGATACCTTACCTATCCTTCAGCTTTTGGATCCATTACA 748  
 234 GlyProAlaHisTrp 238  
 |||||  
 749 GGGCCAGGCCCACTGG 764

seq\_name: gb\_est2:BG703176

seq\_documentation\_block:

LOCUS BG703176 790 bp mRNA linear EST 07-MAY-2001  
 DEFINITION 602686082F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4818405 5',

mRNA sequence.

ACCESSION BG703176

VERSION BG703176.1 GI:13975252

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 790)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10720 row: d column: 22

High quality sequence stop: 766.

FEATURES

source

1..790

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4818405"

/clone.lib="NIH\_MGC\_95"

/tissue\_type="hippocampus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified

pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag

); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',

size-selected for average insert size 2.5 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

BASE COUNT 167 a 228 c 206 g 189 t  
 ORIGIN

Institutes of Health). Note: this is a NIH\_MGC Library."

alignment\_scores:

Quality: 1206.00 Length: 238

Ratio: 5.132 Gaps: 2

Percent Similarity: 98.739 Percent Identity: 97.899

alignment\_block:

US-09-357-675C-21 x BG703176 ..

Align seg 1/1 to: BG703176 from: 1 to: 790

1 MetLeuGlyPheIleThrArgProProHisArgPheLeuSerLeuLeuCy 17  
 |||||  
 79 ATGCTGGGCTTCATCACAGGCTCTCTCAGATTCCTGTCCCTTCGTG 128  
 17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34  
 |||||  
 129 TCTGGACTCCGGATACCTCAACTCTCAGTACTTTGTGCTCAGCCAGGC 178  
 34 roArgAlaMetAlaIleSerSerSerSerCysGluLeuProLeuValAla 50  
 |||||  
 179 CCAGAGCCATGGCTATCTCCTCTCTCTCTCGGAACCTGCCCTGGTGCT 228  
 51 ValCysGlnValThrSerThrProAspLysGlnGlnAsnPhelysthrCy 67  
 |||||  
 229 GTGTGCCAGGTAACATCGAGCCAGACAGCAACAGAACTTTAAACATG 278  
 67 sAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheL 84  
 |||||  
 279 TGTGTAGCTGTTTCGAGAGCTCCAGACTGGGTGCTGCTGCTGGCTTCC 328  
 84 euProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
 |||||  
 329 TGCTTGAGGCATTGACTTCATTGCAGCGGAGCCCTGCAGAGACGCTACAC 378  
 101 LeuSerGluProLeuGlyLysLeuLeuGluGluThrThrGlnLeuAl 117  
 |||||  
 379 CGTCTCTGAACCACTGGGTGGAAACTTTTGGAAAGAAATACACCCAGCTGC 428  
 117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134  
 |||||  
 429 CAGGGAATGGGACTCTGGCTGCTTGGGTGGTTTCCATGACATGGCC 478  
 134 lnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150  
 |||||  
 479 AAGACTGGGAGCAGACTCAGAAAATCTACAATTGTACATGCTGCTGAAC 528  
 151 SerLysGlyAlaValAlaThrTyrArgLysThrHisLeuCysAspVa 167  
 |||||  
 529 AGCAAGGGGCGAGTAGTGCCACTTACAGGAAGACACATCTGTGTGACGT 578  
 167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184  
 |||||  
 579 AGAGATTCCAGGCGAGGGGCTATGTGTGAAGCAACTCTACCATGCCTG 628  
 184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
 |||||  
 629 GGCCAGCTTGTAGTCACCTGTCCAGCACACAGCAGCAAGATTGGTCTA 678  
 201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaG 217  
 |||||  
 679 GCTGTCTGTATTGACATCGGTTCCCTGAACCTCTCTGGCATTTGGCTC 728  
 217 lnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGlySerIleTh 233  
 |||||  
 729 AAGCTGGAGCAGAGATACCTTACCTATCCTTCAGCTTTTGGATCCATTAC 778  
 233 rGlyProAla 236  
 |||||

779 AGGCCAGCC 788

seq\_name: gb\_est1:AL520767

seq\_documentation\_block:

LOCUS AL520767 992 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL520767 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB0021D05 3

prime, mRNA sequence.

ACCESSION AL520767

VERSION AL520767.1 GI:12784260

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 992)

TITLE Li, W. B., Gruber, C., Jessee, J. and Polayes, D.

JOURNAL Full-length cDNA libraries and normalization

COMMENT Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..992

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DB002YD05"

/clone\_lib="LTI\_NFL004\_NBC2"

/sex="male"

/tissue\_type="neuroblastoma cells"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 229 a 267 c 268 g 222 t 6 others

ORIGIN

alignment\_scores:

Quality: 1201.50

Ratio: 5.179 Length: 236

Percent Similarity: 98.305 Gaps: 3

Percent Identity: 96.610

alignment\_block:

US-09-357-675C-21 x AL520767/rev ..

Align seg 1/1 to reverse of: AL520767 from: 1 to: 992

94 AspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLe 110

983 GACCCCTCAGACAGCGTACACCTGCTGACCACTGGTGAAC...TT 937

110 uGluGlu.TyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeu 126

936 TTGRAGRATACACCCAGCTTGCCAGGGAATGTGGACCTGCTGCTCTTG 887

127 GlyGlyPheHisGluArgGlyGlnAspTrpGluClnThrGlnLysIleTy 143

886 GGTGGTTTCATGAGCGTGCCAGACTGGGAGCAGACTCAGAAAAATCTA 837

143 rAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyra 160

836 CAATTGTCAGCTGCTGTACACGAAGGGCAGTAGTGGCCACTTACA 787

160 rgLysThrHisLeuCysAspValcIuileProGlyGlnGlyPrometCys 176

786 GGAAGACACATCTGTGTACGTAGAGATCCAGGCGAGGGCCCTATGTGT 737

177 GluSerAsnSerThrMetProGlyProSerLeuGluSerProValSerTh 193

736 GAAAGCAACTCTACCATGCTGGCCAGCTCTTGAGTCACCTGTACAGCAC 687

193 rProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheProG 210

686 ACCAGCAGGCAAGATTGGTGTAGCTGTCTGTATGACATGCGGTTCCTGT 637

210 luleuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyPro 226

636 AACTCTCTCTGGCATTTGGCTCAAGCTCGACGAGAGATACTTACCTATCCT 587

227 SerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLeuAr 243

586 TCAGCTTTTGGATCCATTACAGGCCCCAGCCACTGGGAGGTGTGCTGCG 537

243 gAlaArgAlaIleGluThrGlnCysTyrValVal.AlAlaAlaGlnCys 259

536 GGCCCGTGTCTATCGAAACCCAGTCTATGTAGTCCMCGCAGCACAGTGT 487

260 GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValas 276

486 GGACGCCACCATGAGAAGAGAGCAAGTTATGCCACACAGCATGGTGTGTA 437

276 pProTrpGlyThrValValAlaArgCysSerGluGlyProGlyLeuCysL 293

436 CCCTTGGGGAACAGTGTGGCCCGCTGCTGTAGGGGCCAGGCCCTGTGCC 387

293 euAlaArqIleAspLeuAsnTyrLeuArgGlnLeuArgArgHisLeuPro 309

386 TTGCCCAATAGACCTCAACTATCTGCACAGTTGCGCCGACACCTGCCT 337

310 ValPheGlnHisArgArgProAspLeuTyrGlyAsnLeuGlyHisProLe 326

336 GTGTTCCAGCAGCCAGCGCTGACCTCTATGGCAATCTGGGTACCCACT 287

326 user 327

286 GTCT 283

seq\_name: gb\_est1:AL522373

seq\_documentation\_block:

LOCUS AL522373 960 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL522373 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB0081M02 3

prime, mRNA sequence.

ACCESSION AL522373

VERSION AL522373.1 GI:12785866

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 960)

TITLE Li, W. B., Gruber, C., Jessee, J. and Polayes, D.

JOURNAL Full-length cDNA libraries and normalization

COMMENT Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..960

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DB0081M02"

/clone\_lib="LTI\_NFL004\_NBC2"

/sex="male"

/tissue\_type="neuroblastoma cells"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 225 a 259 c 255 g 217 t 4 others  
ORIGIN

alignment\_scores:  
Quality: 1179.50 Length: 231  
Ratio: 5.151 Gaps: 2  
Percent Similarity: 99.134 Percent Identity: 98.268

alignment\_block:

US-09-357-675C-21 x AL522373/rev

Align seg 1/1 to reverse of: AL522373 from: 1 to: 960

98 ThrLeuHisLeuSerGluProLeuGlyGlyLysLeuLeu.GluGluTyrT 114  
959 ACCTACACCTGCTCTGAACCACTGGGTGGAAC...TTTGAAGAATACA 913  
114 hrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHis 130  
912 CCCAGCTTGCACGGAATGTGGAATCTGGCTGCTTGGTGGTGGTCCAT 863  
131 GluArgGlyGlnAspTrpGluGlnThrGlnLysIleTyrAsnCysHisVa 147  
862 GACGTGGCCCAAGCACTGGGAGCACTCAGAAATCTACAATTTGACGT 813  
147 lLeuLeuAsnSerLysGlyAlaValAlaThrTyrArgLysThrHisL 164  
812 GCTGCTGAACAGCAAGGGGCACTAGTGGCCACTTACAGGAACACATC 763  
164 euCysaspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180  
762 TGTGTGACGTAGAGATCCAGGCGAGGGGCTATGTGTGAACCAACTCT 713  
181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLY 197  
712 ACCATCGCTGGGCCCCAGTCTGTAGTCACTCAGTCAGCACACACAGCAA 663  
197 sileGlyLeuAlaValCysTyrAspMetArgPheProGluLeuSerLeuA 214  
662 GATTGGTCTAGCTGTCTGCTATGACATGGG...TTCCCTGAACCTCTCTCG 614  
214 lalLeuAlaGlnAlaGlyAlaGluIleLeuThrTyrProSerAlaPheGly 230  
613 CATTTGGCTCAAGCTGGAGAGAGATCTTACCTATCTTCAGCTTTTGA 564  
231 SerIleThrGlyProAlaHisTrpGluValLeuLeuArgAlaAlaI 247  
563 TCCATTACAGGCCCACTTATGGAGGTTGCTGCGGGCCCGTGTAT 514  
247 eGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyArgHisHisG 264  
513 CGAAACCCAGTGTATGTAGTGGCAGCAGCAGATGTGGAGCCACCATG 464  
264 luLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 280  
463 AGAAGAGAGCAAGTTATGGCCACAGCATGTGTGTAGACCCCTGGGAACA 414  
281 ValValAlaArgCysSerGluGlyProGlyLeuLeuAlaAlaAlaAla 297  
413 GTGGTGGCCGCTGCTCTGAGGGGCCAGGCTCTGCTTGGCCGAATAGA 364  
297 pLeuAsnTyrLeuArgGlnLeuArgHisLeuProValPheGlnHisA 314

|||||  
363 CCTCAACTATCTCGCAGAGTTGGCCGACACCTGCTGTGTTCCAGCAC 314  
314 rgArgProaspLeuTyrGlyAsnLeuGlyHisProLeuSer 327  
313 GCAGGGCTGACCTCTTATGCAATCTGGTCACTGGTCACTGTCT 273

seq\_name: gb\_est2:BI757823

seq\_documentation\_block:  
LOCUS BI757823 946 bp mRNA linear EST 25-SEP-2001  
DEFINITION 603030404F1 NIH\_MGC\_114 Homo sapiens cDNA clone IMAGE:5200560 5',  
mRNA sequence.

ACCESSION BI757823  
VERSION BI757823.1 GI:15749401  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 946)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM11502 row: h column: 01

High quality sequence stop: 708.

Location/Qualifiers

1. 946

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5200560"

/clone\_lib="NIH\_MGC\_114"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: EcoRV (destroyed); RNA source anonymous pool of 6

male brains, age range 23-27 yo. Library is oligo-dT

primed and directionally cloned (EcoRV site is destroyed

upon cloning). Average insert size 1.5 kb, insert size

range 1-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber

(Invitrogen). Research Genetics tracking code 019. Note:

this is a NIH\_MGC Library."

BASE COUNT 213 a 264 c 270 g 199 t

ORIGIN

alignment\_scores:

Quality: 1169.00

Ratio: 4.395

Percent Similarity: 91.409

Percent Identity: 83.849

alignment\_block:

US-09-357-675C-21 x BI757823

Align seg 1/1 to: BI757823 from: 1 to: 946

1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCy 17

69 ATGCTGGGCTTCATCACCAGGCTCTCTACAGATTCTGTCCCTTCTGTG 118

17 sProGlyLeuArgIleProGlnLeuSerValLeuCysAlaGlnProArgP 34

119 TCCTGGACTCCGATACCTCAACTCTCAGTACTTGTGTCTCAGCCAGGC 168





```
101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyrThrGlnLeuAl 117
|||||
387 CTGCTGAACCACTGGGTGGAAACCTTTTGAAGAATAACACCCAGCTTGC 436
|||||
117 aArgGluCysGlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyG 134
|||||
437 CAGGGAATGTGGACTCTGGCTGTCCTTGGGTGGTTCATGAGCGTGGCC 486
|||||
134 InAspTrpGluGlnThrGlnLysIleTyrAsnCysHisValLeuLeuAsn 150
|||||
487 AAGACTGGGAGCAGACTCAGAAAATCTACAATTTGTCACTGCTGCTGAAC 536
|||||
151 SerLysGlyAlaValValAlaThrTyrArgLysThrHisLeuCysAspVa 167
|||||
537 AGCAGCGGGGCACTAGTGCCCACTTACAGAGAACACATCTCTGTGACGT 586
|||||
167 lGluIleProGlyGlnGlyProMetCysGluSerAsnSerThrMetProG 184
|||||
587 AGAGATTCAGGCGCGGCGCTATGTGTCAAGCAACTCTACCATGCCGTG 636
|||||
184 lyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200
|||||
637 GGCCCACTGTGAGTCACCTGTGAGCGCAGCAGCAGCAG.ATTGGGCTA 685
|||||
201 AlaValCysTyrAspMetArgPheProGluLeuLeuSerLeuAlaLeuAla 217
|||||
686 GCTGTCTGTATGACGTGGGGTTCCTGGAACCTCTCTGTCATGGCTCA 735
|||||
217 n..AlaGlyAlaGluLeuThrTyrProSerAlaPhe..... 229
|||||
736 GAGCGTGGAGCGGAGATACTTACCTATCCTTTCAGCTTTCGCTCTTGGCG 785
|||||
230 .....GlySerIleThrGlyPro 235
|||||
786 GCCCGCCACTGGAGCGGTGCTGGGGCT 813
|||||

seq_name: gb_est2:B1711300

seq_documentation_block:
LOCUS B1711300 625 bp mRNA linear EST 19-SEP-2001
DEFINITION B196f12.x1 Human insulinoma homo sapiens cDNA 3' similar to
TR:076091 076091 NITRILASE HOMOLOG 1.; mRNA sequence.
ACCESSION B1711300
VERSION B1711300.1 GI:15686995
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence source: 481.
Location/Qualifiers
```

```
source
1. .625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human insulinoma"
/tissue_type="Insulinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
XhoI; Site_2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtm).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

BASE COUNT 139 a 183 c 161 g 142 t
ORIGIN

alignment_scores:
Quality: 1115.00 Length: 208
Ratio: 5.361 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-357-675C-21 x B1711300/rev ..

Align seg 1/1 to reverse of: B1711300 from: 1 to: 625

76 ArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAl 92
|||||
625 AGACTGGGTGCCTGCTGCTTTCCTGCTGAGGCACTTGTACTTCATTGC 576
|||||
92 aArgaspProAlaGluThrLeuHisLeuSerGluProLeuGlyGlyLysL 109
|||||
575 ACGGGACCCCTGCAGAGACGCTACACCTGTCTGAACCACTGGGTGGAAAC 526
|||||
109 euLeuGluGluTyrThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSer 125
|||||
525 TTTTGGAGAATAACACCCAGCTTGCAGGGGATGTGGACTCTGGCTGTCC 476
|||||
126 LeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLysI 142
|||||
475 TTTGGTGGTTTCCATGAGCGTGGCCAAAGACTGGGAGCAGACTCAGAAA 426
|||||
142 eTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValAlaThrT 159
|||||
425 CTACAAATTCACGTGCTGCTGAACAGCAAGGGCGAGTAGTGGCCACTT 376
|||||
159 yTArgLysThrHisLeuCysAspValGluIleProGlyGlnGlyPromet 175
|||||
375 ACAGGGAAGACACATCTGTGTGAGTAGAGATTCCAGGGCAGGGGCTATG 326
|||||
176 CysGluSerAsnSerThrMetProGlyProSerLeuGluSerProValSe 192
|||||
325 TGTGAAAGCAACTCTACCATGCTTGGGCCAGTCTTGTGAGTCACTGTG 276
|||||
192 rThrProAlaGlyLysIleGlyLeuAlaValCysTyrAspMetArgPheP 209
|||||
275 CACACAGCAGCAGACAGATTGGTCTAGCTGTCTGCTATGACATCGCGTCC 226
|||||
209 roGluLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGluIleLeuThrTyr 225
|||||
225 CTGAACTCTCTGGCATTTGGCTCAAGCTGGAGCAGAGATACTTACCTAT 176
|||||
226 ProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeuLe 242
|||||
175 CCTTCAGCTTTTGGATCCATTACAGGCCCCAGGCCACTGGGAGGTGTGCT 126
|||||
242 uArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnC 259
|||||
```

```
125 GCGGGCCCGTGCTATCGAAACCCAGTGTCTATGTAGTGGCAGCAGCACAGT 76
259 ySGlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValVal 275
    |||||
75 GTGGACGCCACCATGAGAGAGAGAGCAAGTTATGCCCCACAGCATGGTGGTA 26
276 AspProTyrGlyThrValValAla 283
    |||||
25 GACCCCTGGGGAAACAGTGGTGGCC 2
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 8, 2002, 10:31:12 ; Search time 2667 Seconds  
(without alignments)  
3568.286 Million cell updates/sec

Title: US-09-357-675C-21  
Perfect score: 327  
Sequence: 1 MLGFTTRPHRFLSLCPGL.....LPVFQHRRLDYLGNLGHPLS 327

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cqn2\_l1/USPFO\_spool/US09357675/runat\_03122002\_142235\_20398/app\_query.fasta\_1.519  
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptlo  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675 -RCGN\_1\_1\_1616\_2runat\_03122002\_142235\_20398 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htg\_mus:\*  
34: em.htg\_pln:\*  
35: em.htg\_rod:\*  
36: em.htg\_mam:\*  
37: em.htg\_vrt:\*  
38: em.sy:\*  
39: em.htgo\_hum:\*  
40: em.htgo\_mus:\*  
41: em.htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	327	100.0	1385	9	AF069987	AF069987 Homo sapi
2	88	26.9	4079	9	AF069984	AF069984 Homo sapi
3	88	26.9	167863	2	AL590651	AL590651 Homo sapi
4	88	26.9	200822	9	AL591806	AL591806 Human DNA
5	66	20.2	847	11	G72919	G72919 MARC 2849-2
6	63	19.3	377	6	AX368386	AX368386 Sequence
7	33	10.1	1338	10	AF069988	AF069988 Mus muscu
8	33	10.1	1365	10	BC021634	BC021634 Mus muscu
9	33	10.1	4481	10	AF069985	AF069985 Mus muscu
10	33	10.1	211772	10	AC084821	AC084821 Mus muscu
11	33	10.1	215043	10	AC087229	AC087229 Mus muscu
12	23	7.0	1214	5	AF284575	AF284575 Xenopus l
13	23	7.0	1214	6	AX419587	AX419587 Sequence
14	18	5.5	115667	2	AC125857	AC125857 Rattus no
15	18	5.5	181583	2	AC105589	AC105589 Rattus no
16	16	4.9	10881	1	AE004129	AE004129 Vibrio ch
17	13	4.0	924	8	AY133544	AY133544 Arabidops
18	13	4.0	1172	8	AF372904	AF372904 Arabidops
19	13	4.0	104386	8	ATT32A17	ATT32A17 Arabidops
20	13	4.0	126375	2	CNS08C81	AL731751 Oryza sat
21	13	4.0	137775	2	CNS08C80	AL731751 Oryza sat
22	13	4.0	199362	8	ATCHRIV24	AL161512 Arabidops
23	12	3.7	1385	3	AF069986	AF069986 Caenorhab
24	12	3.7	169364	3	CEY56A3A	AL132860 Caenorhab
25	11	3.4	1521	3	AF069989	AF069989 Drosophil
26	11	3.4	1576	3	AY089221	AY089221 Drosophil
27	11	3.4	25350	8	SP8C651	AL035570 S.pombe c
28	11	3.4	34013	2	AC010013	AC010013 Drosophil
29	11	3.4	131853	2	AC017755	AC017755 Drosophil
30	11	3.4	168250	3	AC093121	AC093121 Drosophil
31	11	3.4	173497	3	AC093501	AC093501 Drosophil
32	11	3.4	174279	2	AC006169	AC006169 Drosophil
33	11	3.4	186793	2	AC095657	AC095657 Rattus no
34	11	3.4	190353	3	AC105293	AC105293 Drosophil
35	11	3.4	298640	3	AE003467	AE003467 Drosophil
36	10	3.1	10466	1	AE004861	AE004861 Pseudomon
37	10	3.1	15124	1	AE000934	AE000934 Methanoba
38	10	3.1	142756	9	AC005844	AC005844 Homo sapi
39	10	3.1	149000	2	AC129010	AC129010 Leishmani
40	9	2.8	1520	8	AF284572	AF284572 Saccharom
41	9	2.8	4123	1	EAC271724	AF271724 Eubacteri
42	9	2.8	5310	9	HSM803919	AL832609 Homo sapi
43	9	2.8	6597	6	E49884	E49884 Analysis me
44	9	2.8	6597	9	HSCAN	X64228 H.sapiens c
45	9	2.8	6642	9	HUMORF10	D14689 Human mRNA

ALIGNMENTS

RESULT 1



JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)  
 MEDLINE 98337986  
 PUBMED 9671749  
 REFERENCE 2 (bases 1 to 4079)  
 AUTHORS Pekarsky, Y., Campigilio, M., Siprashvili, Z., Druck, T., Sedkov, Y.,  
 Tillib, S., Draganescu, A., Wernuth, P., Rothman, J., Huebner, K.,  
 Buchberg, A.M., Mazo, A., Brenner, C. and Croce, C.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson  
 Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA  
 FEATURES Location/Qualifiers  
 source 1..4079  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1q21-q22"  
 635..3569  
 /gene="NIT1"  
 join(635..680,1288..1383,1636..1890,2011..2114,2321..2454,  
 2609..2734,3001..3569)  
 /gene="NIT1"  
 /product="nitrlase homolog 1"  
 /note="alternatively spliced"  
 635..680  
 /gene="NIT1"  
 /note="alternatively spliced"  
 /number=1  
 join(679..680,1288..1383,1636..1890,2011..2114,2321..2454,  
 2609..2734,3001..3267)  
 /gene="NIT1"  
 /note="NIT1; alternatively spliced"  
 /codon\_start=1  
 /product="nitrlase homolog 1"  
 /protein\_id="AAC39901.1"  
 /db\_xref="GI:3242578"  
 translation="MLGRTPPHRLFLSLCPGLRIPOLSVLCAQPRPRMAISSSS  
 ELPLVAVQVSTDPKQNFKTCAPLVREARAGLAPLPEAFDFIARDPAETLHLS  
 EPLGKLEETQLAREGLMLSGFHERGQDWEQTKIYNCHVLLNSKGAVATYR  
 KTHLCDVIPQQPMCESNTPGPSLESPVSTPAGKTGLAVCYDMRPFELSALAAQ  
 GAELITYPSGSGITGPAHWEVLLRARAIEQCYVVAACGCRHKEKRAVGHSMVVD  
 PWGTVVARSSEGPGLCLARIDLNYLRRLRLPVFQHRPDLNGLHPLS"  
 681..711  
 /gene="NIT1"  
 /note="NIT1; alternatively spliced"  
 712..741  
 /gene="NIT1"  
 /note="NIT1; alternatively spliced"  
 940..1053  
 /gene="NIT1"  
 /note="NIT1; alternatively spliced"  
 1288..1383  
 /gene="NIT1"  
 /number=2  
 1636..1890  
 /gene="NIT1"  
 /number=3  
 2011..2114  
 /gene="NIT1"  
 /number=4  
 2321..2454  
 /gene="NIT1"  
 /number=5  
 2609..2734  
 /gene="NIT1"  
 /number=6  
 3001..3569  
 /gene="NIT1"  
 /number=7  
 940 a 1082 c 1043 g 1012 t 2 others  
 BASE COUNT  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.87e-85 Length: 4079

Score: 88.00 Matches: 88  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.91% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-357-675c-21 (1-327) x AF069984 (1-4079)  
 QY 240 ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnCys 259  
 Db 3001 GTGTGCTCGGGCCGCTGCTATCGAAACCCAGTGTATGTAGTGCAGCAGCACAGTGT 3060  
 QY 260 GlyArgHisHisGlnLysArgAlaSerTyrGlyHisSerMetValValAspProTyrGly 279  
 Db 3061 GGACGCCACCATGAGAAGAGAGCAAGTTATGCCACAGCATGGTGTAGACCCCTGGGA 3120  
 QY 280 ThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn 299  
 Db 3121 ACAGTGGTGGCCGCTGCTCGAGGGCCAGGCCTCTGCCTTCCCGGAATAGACCTCAAC 3180  
 QY 300 TyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyr 319  
 Db 3181 TATCTCGACAGTTGCCCGCACACCTGCCTGTGTTCAGCACCGCAGCCCTGACCTCTAT 3240  
 QY 320 GlyAsnLeuGlyHisProLeuSer 327  
 Db 3241 GCAATCTGGGTCAACCACTGTCT 3264  
 RESULT 3  
 AL590651/c  
 LOCUS 167863 bp DNA linear HTG 18-AUG-2001  
 DEFINITION Homo sapiens chromosome 1 clone RP11-137A12, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 9 unordered pieces.  
 ACCESSION AL590651  
 VERSION AL590651.4 GI:13990236  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 167863)  
 AUTHORS Harrison, E.  
 DIRECT SUBMISSION  
 TITLE Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 JOURNAL requests: clonerequest@sanger.ac.uk  
 COMMENT On May 7, 2001 this sequence version replaced gi:13751001.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 ----- Project information  
 Center project name: bA137A12  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 164729 bases at least Q40  
 Consensus quality: 165486 bases at least Q30  
 Consensus quality: 166169 bases at least Q20  
 Insert size: 167063; sum-of-contigs  
 Insert size: 171878; 3.4% error; agarose-fp  
 Quality coverage: 8.28x in Q20 bases; sum-of-contigs Quality  
 coverage: 8.11x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will

\* be preserved.  
\* 1 9373: contig of 9373 bp in length  
\* 9374 9473: gap of 100 bp  
\* 9474 29891: contig of 20418 bp in length  
\* 29892 29991: gap of 100 bp  
\* 29992 36239: contig of 6248 bp in length  
\* 36240 36339: gap of 100 bp  
\* 36340 103784: contig of 67445 bp in length  
\* 103785 103884: gap of 100 bp  
\* 103885 114878: contig of 10994 bp in length  
\* 114879 114978: gap of 100 bp  
\* 114979 126347: contig of 11369 bp in length  
\* 126348 126447: gap of 100 bp  
\* 126448 137764: contig of 11317 bp in length  
\* 137765 137864: gap of 100 bp  
\* 137865 163160: contig of 25296 bp in length  
\* 163161 163260: gap of 100 bp  
\* 163261 167863: contig of 4603 bp in length.

## FEATURES

## source

1..167863  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone\_lib="RPCI-11.1"  
/clone\_lib="RPCI-11.1"

## misc\_feature

1..9373  
/note="assembly\_fragment:02360  
fragment\_chain:1  
clone\_end:17  
vector\_side:left"

## misc\_feature

9474..29891  
/note="assembly\_fragment:01854  
fragment\_chain:1"

## misc\_feature

29992..36239  
/note="assembly\_fragment:00805  
fragment\_chain:1"

## misc\_feature

36340..103784  
/note="assembly\_fragment:01467  
fragment\_chain:1"

## misc\_feature

103885..114878  
/note="assembly\_fragment:00991  
fragment\_chain:2"

## misc\_feature

114979..126347  
/note="assembly\_fragment:01503  
fragment\_chain:2"

## misc\_feature

126448..137764  
/note="assembly\_fragment:01297  
fragment\_chain:2"

## misc\_feature

137865..163160  
/note="assembly\_fragment:00117  
fragment\_chain:2"

## misc\_feature

163261..167863  
/note="assembly\_fragment:01267  
fragment\_chain:2  
clone\_end:SP6  
vector\_side:right"

BASE COUNT 43927 a 39935 c 40002 g 43197 t 802 others  
ORIGIN

## Alignment Scores:

Pred. No.: 1.98e-83 Length: 167863  
Score: 88.00 Matches: 88  
Percent Similarity: 100.00% Conservatived: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.91% Indels: 0  
DB: 2 Gaps: 0

US-09-357-675c-21 (1-327) x AL590651 (1-167863)

Qy 240 ValLeuArgAlaArgAlaLeGluThrGlnCysTyrValValAlaAlaGlnCys 259

|||||  
Db 149680 GGTGTCGCGCGCGCTATCGAACCAGTCTATGATGGCAGCAGCAGTGT 149621

Qy 260 GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGly 279  
|||||  
Db 149620 GGACCCACCATGAGAGAGACCAAGTTATGCCACAGCATGGTAGACCCCTGGGA 149561  
|||||  
Qy 280 ThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn 299  
|||||  
Db 149560 ACAGTGGTGGCCGCTCTCTGAGGGCCAGCGCTCTGCCTTGCCCGAATAGACCTCAAC 149501  
|||||  
Qy 300 TyrLeuArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyr 319  
|||||  
Db 149500 TATCTGCGACAGTTGCGCGACACCTGCTGTCTCCAGCACCGAGGCTGACCTCTAT 149441  
|||||  
Qy 320 GlyAsnLeuGlyHisProLeuSer 327  
|||||  
Db 149440 GCAATCTGGGTCAACCCTGCT 149417  
|||||

## RESULT 4

## AL591806

## LOCUS

DEFINITION Human DNA sequence from clone RP11-544M22 on chromosome 1, complete sequence.

## ACCESSION

## AL591806

## VERSION

AL591806.16 GI:18476709

## KEYWORDS

HTG.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Feb 1, 2002 this sequence version replaced gi:17902927.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr1  
RP11-544M22 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6  
This sequence is the entire insert of clone RP11-544M22 The true left end of clone RP11-137A12 is at 156538 in this sequence. The true right end of clone RP11-381D2 is at 145015 in this sequence.

## FEATURES

## source

1..200822  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-544M22"  
/clone\_lib="RPCI-11.2"  
10395..10397  
/note="Sequence from overlapping clone RP11-381D2



misc_feature	misc_feature	misc_feature	misc_feature
(AL162592)..Assembly confirmed by restriction digest. 440398..44107	/note="Sequence from overlapping clone RP11-381D2 (AL162592)..Assembly confirmed by restriction digest." 59595..59604	/note="Sequence from overlapping clone RP11-381D2 (AL162592)..Assembly confirmed by restriction digest." 90797..90883	/note="Single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly." 90861..90883
			/note="Sequence from uni-directional dgpp big dye terminator reads only." a 49052 c 46183 q 50372 t
			BASE COUNT 55215 a 49052 c 46183 q 50372 t

Primer B: GCTGAGCCTGTGTTTCATCA  
STS size: 700  
PCR Profile:  
Hotstart: 95 degrees for 15 minutes  
Denature: 95 degrees for 30 seconds  
Anneal: 58 degrees  
Extension: 68 degrees for 2 minutes  
Cycles: 32 to 45

Protocol:  
Template: 50-200 ng genomic DNA  
Primer: each 20 pmoles  
dNTPs: each 88 uM  
Taq Polymerase: 0.25 units (Qiagen HotStar)

Buffer:  
Commercially supplied Qiagen HotStar buffer

The STS is derived from PCR amplicons generated from genomic DNA, sequenced from each end using the amplification primers. The sequence does not necessarily represent the entire amplicon. Sequence derived from PolyPhred was trimmed from each end of each unique contig until five consecutive bases exceeded a quality score threshold of 20, and the next 10 bases averaged a quality score of 20 or greater. Amplicon size was estimated by agarose gel electrophoresis.

FEATURES  
source Location/Qualifiers  
1..847  
/organism="Sus scrofa"  
/strain="white composite, duroc, meishan, minzhu, fengjing, crossbreds"  
/db\_xref="taxon:9823"  
/sex="male and female"  
/clone\_lib="SCF - porcine spleen"  
/dev\_stage="adult"  
/note="Organ: spleen"  
<1..847

STS 205 a 249 c 200 g 193 t  
BASE COUNT  
ORIGIN

Alignment Scores:  
Pred. No.: 1..3e-61 Length: 847  
Score: 66.00 Matches: 66  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 20.18% Indels: 0  
DB: 11 Gaps: 0

US-09-357-675C-21 (1-327) x G72919 (1-847)

Qy	240	ValLeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCys	259
Db	484	GTGTGTCTGCGGGCCCGAGCCATTGAACACTCAGTGTCTATGTAGTGGCGGACACAGTGT	543
Qy	260	GlyArgHisHisGluLysArgAlaSerTyrGlyHisSerMetValValAspProTrrpGly	279
Db	544	GGACCCACCATGAGAGAGAGCAAGTATTGGCCATAGCATGTGTGTAGATCCCTGGGGA	603
Qy	280	ThrValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsn	299
Db	604	ACGTGTGTGCGCCCTGCTCTGAAGGACACAGGCGCTGCCTTGGCCGAATTGACCTCAAT	663
Qy	300	TyrLeuArgGlnLeuArg	305
Db	664	TATCTGCACAGTTGCCG	681

RESULT 6  
AX368386/c  
LOCUS AX368386 377 bp DNA linear PAT 16-FEB-2002  
DEFINITION Sequence 1096 from Patent WO0204514.  
ACCESSION AX368386  
VERSION AX368386.1 GI:18956459  
KEYWORDS human.  
SOURCE

JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	95	(15),	8744-8749	(1998)
MEDLINE	98337986				
PUBMED	9671749				
REFERENCE	2	(bases 1 to 1338)			
AUTHORS	Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y., Tillib,S., Draganescu,A., Wernuth,P., Rothman,J.J., Huebner,K., Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA				
FEATURES	Location/Qualifiers				
source	1..1338	/organism="Mus musculus"			
		/db_xref="taxon:10090"			
		/chromosome="1"			
		/map="1q21-q23"			
gene	1..1338	/gene="Nt1l"			
CDS	58..1029	/gene="Nt1l"			
		/codon_start=1			
		/product="nitrlase 1"			
		/protein_id="AAC40185.1"			
		/db_xref="GI:3228668"			
		/translation="MLGFITRPHQLLCTGYRLRTPVLCQPRPRTSSSTSWELPL VAVQVTSTPNKQENFKTCAELVOEARLGAFLAPEAFDFIARNPAETLLLESEPL GDLLGOXSOLARECGIWLSGLGPHFERGDMEONKIYCNHVLLNSGSVVASVRKTHL CDVEIPGGQPMRESNYTKPGGTLEPPVKTPAGKVGLAICYDMRFELSCLKLAAGAEL LTVSFASGVTFPAHWEVLRLARAIESQCIVIAACQGRHHETHASTRGHSVMVDPMGT VVARCEGFGCLARLDHLFLOQMROHLPVFQHRRPDLXGSLGHPLS"			
BASE COUNT	347 a	335 c	330 g	326 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	1.36e-25	Length:	1338		
Score:	33.00	Matches:	33		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	10.09%	Indels:	0		
DB:	10	Gaps:	0		
US-09-357-675C-21	(1-327) x AF069988	(1-1338)			
QY	266	ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValalaArgCys	285		
Db	841	AGAGCAAGTTATGGCCATAGCATGGTGTTGACCGTGGGCACAGTGGTGGCCCGCTGC	900		
QY	286	SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu	298		
Db	901	TCCGAGGGACCAGGCCCTCTGCTTCGTAATGTATCTC	939		
RESULT	8				
LOCUS	BC021634	Mus musculus, nitrlase 1, clone MGC:13825 IMAGE:4008543, mRNA,	1365 bp	mRNA	linear
DEFINITION		complete cds.			
ACCESSION	BC021634				
VERSION	BC021634.1	GI:18204912			
KEYWORDS	MGC.				
SOURCE		house mouse.			
ORGANISM		Mus musculus			
REFERENCE		Eukaryota; Metazoa; Chordata; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS		1 (bases 1 to 1365)			
TITLE		Straussberg,R.			
JOURNAL		Direct Submission			
		Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK		NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT		Contact: MGC help desk Email: cgapbs-r@mail.nih.gov			

Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
 http://www.systemsbio.org  
 contact: amadan@systemsbiology.org  
 Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAP Plate: 18 Row: e Column: 20  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6754855.

## FEATURES

## Location/Qualifiers

## source

1. 1365

/organism="Mus musculus"

/db\_xref="taxon:10090"

/map="CZECH II"

/clone="MGC:13825 IMAGE:4008543"

/tissue\_type="Mammary tumor metastasized to lung."

MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR

enhancer."

/clone.lib="NCI\_CGAP\_Lu30"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

59. 1030

/codon\_start=1

/product="nitrlase 1"

/protein\_id="AAH21634.1"

/db\_xref="GI:18204913"

/db\_xref="LocusID:27045"

/translation="MLGFITRPPHQLLCTGYRLRLTPVLCQPRPTMSSSTSWELPL

VAVCQVSTPNKQENFTCAELVQEARLGLAFLPEAFDFIARNPAETLLSEPLN

GDLLGOXSOLARECGIWLGLGFHGERGDWQONOKIYNCHVLLNSKGSVASYRKTLL

CDVEIPQGGPMRESNYTKPGTLEPPVKTPAGKGLAICYDMRPELSLKLQAQAEI

LTYSARGSVTPGAHWEVLLRARAIESQCYVIAAQCGRHHETRASVGHSMVDPWGT

VVARSCEGPGCLARIDLHFLQQRHQLPVPFQHRPDDLYSLGHPLS"

BASE COUNT 368 a 334 c 331 g 332 t

## ORIGIN

## Alignment Scores:

Pred. No.: 1.39e-25 Length: 1365  
 Score: 33.00 Matches: 33  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.09% Indels: 0  
 DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x BC021634 (1-1365)

QY 266 ArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThrValValAlaArgCys 285

Db 842 AGAGCAAGTTATGGCCATAGCATGTGGTTGACCGTGGGGACACAGTGGTCCCGCTGC 901

QY 286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298

Db 902 TCCGAGGGACACAGGCTCTGCTTGTCTCGAATTGATCTC 940

## RESULT 9

## AF069985

## LOCUS

AF069985 4481 bp DNA linear ROD 23-JUL-1998

DEFINITION Mus musculus nitrlase homolog 1 (Nt1) gene, alternatively spliced

product, complete cds.

ACCESSION AF069985

VERSION AF069985.1 GI:3242979

KEYWORDS

SOURCE Mus musculus.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4481)

AUTHORS Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Wermuth,P., Rothman,J.H., Huebner,K.,  
 Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Nitrlase and Fhit homologs are encoded as fusion proteins in

Drosophila melanogaster and Caenorhabditis elegans

Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8744-8749 (1998)

98337986

9671749

2 (bases 1 to 4481)

Pekarsky,Y., Campiglio,M., Siprashvili,Z., Druck,T., Sedkov,Y.,

Tillib,S., Draganescu,A., Wermuth,P., Rothman,J., Huebner,K.,

Buchberg,A.M., Mazo,A., Brenner,C. and Croce,C.M.

Direct Submission

Submitted (03-JUN-1998) Kimmel Cancer Inst., Thomas Jefferson

Univ., 233 S. Tenth St., Philadelphia, PA 19107, USA

## FEATURES

## Location/Qualifiers

## source

1. 4481

/organism="Mus musculus"

/db\_xref="taxon:10090"

/map="1q21-q23"

541. 3944

/gene="Nt1"

join(541. 606,1177. 1263,1638. 1889,2015. 2118,2362. 2495,

2626. 2751,3392. 3944)

/gene="Nt1"

/product="nitrlase homolog 1"

/note="alternatively spliced"

541. 606

/gene="Nt1"

/note="alternatively spliced"

/number=1

join(605. 606,1177. 1263,1638. 1889,2015. 2118,2362. 2495,

2626. 2751,3392. 3658)

/gene="Nt1"

/note="alternatively spliced"

/codon\_start=1

/product="nitrlase homolog 1"

/protein\_id="AAC40184.1"

/db\_xref="GI:3242980"

/translation="MLGFITRPPHQLLCTGYRLRLTPVLCQPRPTMSSSTSWELPL

VAVCQVSTPNKQENFTCAELVQEARLGLAFLPEAFDFIARNPAETLLSEPLN

GDLLGOXSOLARECGIWLGLGFHGERGDWQONOKIYNCHVLLNSKGSVASYRKTLL

CDVEIPQGGPMRESNYTKPGTLEPPVKTPAGKGLAICYDMRPELSLKLQAQAEI

LTYSARGSVTPGAHWEVLLRARAIESQCYVIAAQCGRHHETRASVGHSMVDPWGT

VVARSCEGPGCLARIDLHFLQQRHQLPVPFQHRPDDLYSLGHPLS"

863. 986

/gene="Nt1"

/note="alternatively spliced"

/number=2

1177. 1263

/gene="Nt1"

/number=3

1638. 1889

/gene="Nt1"

/number=4

2015. 2118

/gene="Nt1"

/number=5

2362. 2495

/gene="Nt1"

/number=6

2626. 2751

/gene="Nt1"

/number=7

3392. 3944

/gene="Nt1"

/number=8

BASE COUNT 1101 a 1132 c 1133 g 1115 t

## ORIGIN

## Alignment Scores:

Pred. No.: 4.07e-25 Length: 4481

Score: 33.00 Matches: 33

```

Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 10.09%   Indels: 0
DB: 10   Gaps: 0

US-09-357-675C-21 (1-327) x AF069985 (1-4481)

QY 266 ArgAlaSerTyrGlyHisSerMetValValAspProTrrpGlyThrValValalaArgCys 285
Db 3470 AGAGCAAGTTATGGCCATAGCATGTTGACCGTGGGGCACAGTGGTGGCCCGCTGC 3529

QY 286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298
Db 3530 FCCGAGGACCGAGGCTCTGCTTCTCGAATTGATCTC 3568

RESULT 10
AC084821
LOCUS AC084821 211772 bp DNA linear ROD 19-FEB-2002
DEFINITION Mus musculus chromosome 1 clone rp23-395h6, complete sequence.
ACCESSION AC084821
VERSION AC084821.25 GI:18702388
KEYWORDS HTG.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL Deschamps, S., Gu, W. and Roe, B. A.
PUBLISHED 1 (bases 1 to 211772)
JOURNAL Mus musculus BAC Clone rp23-395h6
REFERENCE
AUTHORS Deschamps, S., Gu, W. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Deschamps, S., Gu, W. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Feb 19, 2002 this sequence version replaced gi:18390260.
-----
FEATURES             Location/Qualifiers
     source           1. .211772
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="rp23-395h6"
BASE COUNT 55607 a 48743 c 49203 g 58215 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 1.33e-23 Length: 211772
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.09% Indels: 0
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x AC084821 (1-211772)
QY 266 ArgAlaSerTyrGlyHisSerMetValValAspProTrrpGlyThrValValalaArgCys 285
Db 2809 AGAGCAAGTTATGGCCATAGCATGTTGACCGTGGGGCACAGTGGTGGCCCGCTGC 2868

QY 286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298

```

```

Db 2869 TCCGAGGACCGAGGCTCTGCTTCTCGAATTGATCTC 2907

RESULT 11
AC087229/c
LOCUS AC087229 215043 bp DNA linear ROD 14-AUG-2002
DEFINITION Mus musculus chromosome 1 clone rp23-191a19 strain C57BL/6J,
complete sequence.
ACCESSION AC087229
VERSION AC087229.24 GI:22094398
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL Jiang, X., Song, L., Gu, W. and Roe, B. A.
PUBLISHED 1 (bases 1 to 215043)
JOURNAL Mus musculus Chromosome 1 BAC Clone rp23-191a19
REFERENCE
AUTHORS Jiang, X., Song, L., Gu, W. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, X., Song, L., Gu, W. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, X., Song, L., Gu, W. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, X., Song, L., Gu, W. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, X., Song, L., Gu, W. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 3, 2002 this sequence version replaced gi:22038565.
-----
FEATURES             Location/Qualifiers
     source           1. .215043
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="rp23-191a19"
BASE COUNT 55806 a 49340 c 51162 g 58732 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 1.35e-23 Length: 215043
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.09% Indels: 0
DB: 10 Gaps: 0

```

US-09-357-675C-21 (1-327) x AC087229 (1-215043)

QY 266 ArgAlaSerTyrGlyHisSerMetValValAspProTrrpGlyThrValValAlaArgCys 285  
 DB 88091 AGAGCAAGTATGGCCATGACATGATGGTTGACCGTGGGGCAGAGTGGTGGCCCGCTGC 88032  
 QY 286 SerGluGlyProGlyLeuCysLeuAlaArgIleAspLeu 298  
 DB 88031 TCCGAGGGACAGGCTCTGCTTGCATGATGATCTC 87993

RESULT 12  
 AF284575  
 LOCUS 1214 bp mRNA linear VRT 22-JUL-2000  
 DEFINITION Xenopus laevis Nit protein 1 (Nit1) mRNA, complete cds.  
 ACCESSION AF284575  
 VERSION AF284575.1 GI:9367117  
 KEYWORDS  
 SOURCE Xenopus laevis.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 Xenopodinae; Xenopus.  
 REFERENCE  
 AUTHORS Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,  
 Bieganski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.  
 TITLE Crystal structure of the worm Nit1 Rosetta Stone protein reveals  
 a Nit tetramer binding two Fhit dimers  
 JOURNAL Curr. Biol. 10 (15), 907-917 (2000)  
 MEDLINE 20414396  
 PUBMED 10959838  
 REFERENCE  
 AUTHORS Pace,H.C., Hodawadekar,S.C., Draganescu,A., Huang,J.,  
 Bieganski,P., Pekarsky,Y., Croce,C.M. and Brenner,C.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JUL-2000) Kimmel Cancer Center, Thomas Jefferson  
 University, 233 S 10th Street, Rm. 826, Philadelphia, PA 19107, USA  
 FEATURES  
 source  
 1. .1214  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 gene  
 1. .1214  
 /gene="Nit1"  
 CDS  
 180..1046  
 /gene="Nit1"  
 /note="Nit1: similar to the Nit domains of Caenorhabditis  
 elegans NitFhit and Drosophila melanogaster NitFhit"  
 /codon\_start=1  
 /product="Nit protein 1"  
 /protein\_id="AAF87104.1"  
 /db\_xref="GI:9367118"

BASE COUNT 281 a 305 c 345 g 282 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9.07e-15 Length: 1214  
 Score: 23.00 Matches: 23  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.03% Indels: 0  
 DB: 5 Gaps: 0

US-09-357-675C-21 (1-327) x AF284575 (1-1214)

QY 236 AlaHisTrpGluValLeuLeuArgAlaArgAlaGluThrGlnCysTyrValValAla 255  
 DB 765 GCACATTGGGAGGTGTGTGTGAGAGCCCGTGCATAGAAACCCAGTGTACGTAGTTGCA 824

QY 256 AlaAlaGln 258  
 DB 825 GCGGCACAG 833  
 RESULT 13  
 AX419587  
 LOCUS 115667 bp DNA linear PAT 18-JUN-2002  
 DEFINITION Sequence 10 from Patent WO0187958.  
 ACCESSION AX419587  
 VERSION AX419587.1 GI:21523970  
 KEYWORDS  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 Xenopodinae; Xenopus.  
 REFERENCE  
 AUTHORS Croce,C., Brenner,C. and Pekarski,Y.  
 TITLE Crystal structure of worm nitfhit reveals that a nit tetramer binds  
 two fhit dimers  
 JOURNAL Patent: WO 0187958-A 10 22-NOV-2001;  
 Thomas Jefferson University (US)  
 FEATURES  
 source  
 1. .1214  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 BASE COUNT 281 a 305 c 345 g 282 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.07e-15 Length: 1214  
 Score: 23.00 Matches: 23  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.03% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-357-675C-21 (1-327) x AX419587 (1-1214)  
 QY 236 AlaHisTrpGluValLeuLeuArgAlaArgAlaGluThrGlnCysTyrValValAla 255  
 DB 765 GCACATTGGGAGGTGTGTGTGAGAGCCCGTGCATAGAAACCCAGTGTACGTAGTTGCA 824  
 QY 256 AlaAlaGln 258  
 DB 825 GCGGCACAG 833  
 RESULT 14  
 AC125857/c  
 LOCUS 115667 bp DNA linear HTG 23-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-421M5, \*\*\* SEQUENCING IN PROGRESS  
 ACCESSION AC125857  
 VERSION AC125857.1 GI:21671452  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Rattus norvegicus.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 115667)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

```

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 115667)
Worley,K.C.

Direct Submission
Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 115667)
Worley,K.C.

Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWUD
Center clone name: CH230-421M5
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 72804 bases at least Q40
Consensus quality: 72778 bases at least Q30
Consensus quality: 80916 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1519: contig of 1519 bp in length
* 1520 1619: gap of unknown length
* 1620 3041: contig of 1422 bp in length
* 3042 3141: gap of unknown length

```

```

3142 4192: contig of 1051 bp in length
4193 4292: gap of unknown length
4293 6074: contig of 1782 bp in length
6075 6174: gap of unknown length
6175 7346: contig of 1172 bp in length
7347 7446: gap of unknown length
7447 8693: contig of 1247 bp in length
8694 8793: gap of unknown length
8794 10240: contig of 1447 bp in length
10241 10340: gap of unknown length
10341 12235: contig of 1895 bp in length
12236 12335: gap of unknown length
12337 13390: contig of 1055 bp in length
13391 13490: gap of unknown length
13491 14571: contig of 1081 bp in length
14572 14671: gap of unknown length
14672 15864: contig of 1193 bp in length
15865 15964: gap of unknown length
15965 17316: contig of 1352 bp in length
17317 17416: gap of unknown length
17417 18566: contig of 1240 bp in length
18567 18757: gap of unknown length
18758 20056: contig of 1300 bp in length
20057 20156: gap of unknown length
20157 21931: contig of 1775 bp in length
21932 22031: gap of unknown length
22032 24000: contig of 1969 bp in length
24001 24100: gap of unknown length
24101 28222: contig of 2122 bp in length
28223 28322: gap of unknown length
28323 27382: contig of 1060 bp in length
27383 27482: gap of unknown length
27483 28666: contig of 1184 bp in length
28667 28766: gap of unknown length
28767 31151: contig of 2385 bp in length
31152 32151: gap of unknown length
32152 32815: contig of 1564 bp in length
32816 32915: gap of unknown length
32916 34953: contig of 2038 bp in length
34954 35053: gap of unknown length
35054 36996: contig of 1643 bp in length
36997 37996: gap of unknown length
37997 39079: contig of 2283 bp in length
39080 39179: gap of unknown length
39180 41478: contig of 2299 bp in length
41479 41578: gap of unknown length
41579 43648: contig of 2070 bp in length
43649 43748: gap of unknown length
43749 45639: contig of 1891 bp in length
45640 45739: gap of unknown length
45740 47441: contig of 1702 bp in length
47442 47541: gap of unknown length
47542 49632: contig of 2091 bp in length
49633 49732: gap of unknown length
49733 52604: contig of 2872 bp in length
52605 52704: gap of unknown length
52705 55288: contig of 3584 bp in length
55289 56388: gap of unknown length
56389 58558: contig of 2170 bp in length
58559 58658: gap of unknown length
58659 61781: contig of 3123 bp in length
61782 61881: gap of unknown length
61882 64670: contig of 2789 bp in length
64671 64770: gap of unknown length
64771 69117: contig of 4347 bp in length
69118 69217: gap of unknown length
69218 72604: contig of 3387 bp in length
72605 72704: gap of unknown length
72705 76341: contig of 3637 bp in length
76342 76441: gap of unknown length
76442 80834: contig of 4393 bp in length
80835 80934: gap of unknown length
80935 84092: contig of 3158 bp in length

```

\* 84093 84192: gap of unknown length  
 \* 84193 89965: contig of 5773 bp in length  
 \* 89966 90065: gap of unknown length  
 \* 90066 93596: contig of 3531 bp in length  
 \* 93597 93696: gap of unknown length  
 \* 93697 99334: contig of 6238 bp in length  
 \* 99335 100034: gap of unknown length  
 \* 100035 106684: contig of 6650 bp in length  
 \* 106685 106784: gap of unknown length  
 \* 106785 115667: contig of 8883 bp in length.

## FEATURES

source

1..115667  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-421M5"  
 BASE COUNT 29660 a 25606 c 26021 g 29956 t 4424 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.51e-07 Length: 115667  
 Score: 18.00 Matches: 18  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.50% Indels: 0  
 DB: 2 Gaps: 0

US-09-357-675C-21 (1-327) x AC125857 (1-115667)

QY 76 ArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArg 93

Db 36950 AGACTGGCGCTTGGCTGCTTCTGCTGAGCATTTGACTTTATTGCACGA 36897

## RESULT 15

AC105589

LOCUS Rattus norvegicus clone CH230-242B2, linear HTG 13-JUL-2002

DEFINITION \*\*\* 67 unordered pieces. \*\*\* SEQUENCING IN PROGRESS

AC105589

AC105589.2 GI:21736456

VERSION HTG: HTGS-PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 181583)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Ayale, M., Banks,T., Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Tang,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Thomas,S., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

## TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

\* 14068 14167: gap of unknown length  
\* 14168 15370: contig of 1203 bp in length  
\* 15371 15470: gap of unknown length  
\* 15471 16926: contig of 1456 bp in length  
\* 16927 17026: gap of unknown length  
\* 17027 18088: contig of 1062 bp in length  
\* 18089 18188: gap of unknown length  
\* 18189 19315: contig of 1127 bp in length  
\* 19316 19415: gap of unknown length  
\* 19416 20772: contig of 1357 bp in length  
\* 20773 20872: gap of unknown length  
\* 20873 21937: contig of 1065 bp in length  
\* 21938 22037: gap of unknown length  
\* 22038 23147: contig of 1110 bp in length  
\* 23148 23247: gap of unknown length  
\* 23248 24803: contig of 1556 bp in length  
\* 24804 24903: gap of unknown length  
\* 24904 26111: contig of 1208 bp in length  
\* 26112 26211: gap of unknown length  
\* 26212 27647: contig of 1436 bp in length  
\* 27648 27747: gap of unknown length  
\* 27748 30257: contig of 2510 bp in length  
\* 30258 30357: gap of unknown length  
\* 30358 31725: contig of 1368 bp in length  
\* 31726 31825: gap of unknown length  
\* 31826 32912: contig of 1087 bp in length  
\* 32913 33012: gap of unknown length  
\* 33013 34593: contig of 1581 bp in length  
\* 34594 34693: gap of unknown length  
\* 34694 36338: contig of 1845 bp in length  
\* 36339 36438: gap of unknown length  
\* 36439 38139: contig of 1701 bp in length  
\* 38140 38239: gap of unknown length  
\* 38240 39349: contig of 1110 bp in length  
\* 39350 39449: gap of unknown length  
\* 39450 40968: contig of 1519 bp in length  
\* 40969 41068: gap of unknown length  
\* 41069 42068: contig of 1000 bp in length  
\* 42069 42168: gap of unknown length  
\* 42169 43869: contig of 1701 bp in length  
\* 43870 43969: gap of unknown length  
\* 43970 46141: contig of 2172 bp in length  
\* 46142 46241: gap of unknown length  
\* 46242 48158: contig of 1917 bp in length  
\* 48159 48258: gap of unknown length  
\* 48259 50613: contig of 2355 bp in length  
\* 50614 50713: gap of unknown length  
\* 50714 52725: contig of 2012 bp in length  
\* 52726 52825: gap of unknown length  
\* 52826 54534: contig of 1709 bp in length  
\* 54535 54634: gap of unknown length  
\* 54635 56455: contig of 1821 bp in length  
\* 56456 58555: gap of unknown length  
\* 58556 58494: contig of 1939 bp in length  
\* 58594 58594: gap of unknown length  
\* 58595 60496: contig of 1902 bp in length  
\* 60497 60596: gap of unknown length  
\* 60597 62628: contig of 2032 bp in length  
\* 62629 62728: gap of unknown length  
\* 62729 65060: contig of 2332 bp in length  
\* 65061 65160: gap of unknown length  
\* 65161 66543: contig of 1383 bp in length  
\* 66544 66643: gap of unknown length  
\* 66644 68412: contig of 1769 bp in length  
\* 68413 68512: gap of unknown length  
\* 68513 71585: contig of 3073 bp in length  
\* 71586 71685: gap of unknown length  
\* 71686 74575: contig of 2890 bp in length  
\* 74576 74675: gap of unknown length  
\* 74676 77865: contig of 3190 bp in length  
\* 77866 77965: gap of unknown length  
\* 80442: contig of 2477 bp in length  
\* 80443 80542: gap of unknown length

\* 80543 83108: contig of 2566 bp in length  
\* 83109 83208: gap of unknown length  
\* 83209 86895: contig of 3687 bp in length  
\* 86896 86995: gap of unknown length  
\* 86996 91038: contig of 4043 bp in length  
\* 91039 91138: gap of unknown length  
\* 91139 94412: contig of 3274 bp in length  
\* 94413 94512: gap of unknown length  
\* 94513 96959: contig of 2447 bp in length  
\* 96960 97059: gap of unknown length  
\* 97060 98988: contig of 1929 bp in length  
\* 98989 99088: gap of unknown length  
\* 99089 101633: contig of 2545 bp in length  
\* 101634 101733: gap of unknown length  
\* 101734 105899: contig of 4166 bp in length

Alignment Scores:

Pred. No.: 2.27e-07 Length: 181583  
Score: 18.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.50% Indels: 0  
DB: 2 Gaps: 0

US-09-357-675C-21 (1-327) x AC105589 (1-181583)

Oy 76 ArgLeuGlyAlaCysLeuAlaPheLeuProGluAlaPheAspPheIleAlaArg 93

Db 133573 AGACGCGCGCTTGCCTTCTGCTGAGGCATTGACTTTATTGACGA 133626

Search completed: December 8, 2002, 11:33:16

Job time : 2957 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 8, 2002, 10:30:41 ; Search time 255 seconds  
(without alignments)  
2887.857 Million cell updates/sec

Title: US-09-357-675C-21  
Perfect score: 327  
Sequence: 1 MLGFTTPRHFLSLCPGL.....LPVFQHRRLDGLNGLHPLS 327

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4368727

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+D2n model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09357675/runat\_03122002\_142235\_20387/app\_query.fasta\_1.519  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=huma40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09357675 -SCGN\_1\_1\_125 -runat\_03122002\_142235\_20387 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*  
1: /SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	326	99.7	1416	21	Nucleotide sequence
2	253	77.4	1382	21	Human prostate can
3	153	46.8	1203	20	DNA encoding a hum
4	63	19.3	377	24	cDNA encoding lung
5	23	7.0	1214	24	Xenopus laevis N1t
6	18	5.5	539	22	Human immune/haema
7	13	4.0	604	24	Arabidopsis thalia
8	11	3.4	1495	23	Drosophila melanog
9	11	3.4	3548	23	Drosophila melanog
10	11	3.4	5692	23	Drosophila melanog
11	9	2.8	242	24	Human ORFX polynuc
12	9	2.8	6597	22	Human polynucleoti
13	9	2.8	6597	24	Human cDNA differe
14	9	2.8	69936	21	N. meningitidis pa
15	9	2.8	34980	21	Neisseria meningit
16	9	2.8	1437668	21	N. meningitidis B
17	8	2.4	44	24	PCR primer #7 used
18	8	2.4	56	21	Human BPI PCR prim
19	8	2.4	90	21	Human secreted pro
20	8	2.4	115	22	Escherichia coli J
21	8	2.4	115	24	Synthetic DNA targ
22	8	2.4	250	23	Genomic sequence #
23	8	2.4	280	22	Human nervous syst
24	8	2.4	291	21	Human secreted pro
25	8	2.4	314	24	Human ORFX polynuc
26	8	2.4	330	23	Drosophila melanog
27	8	2.4	351	22	Escherichia coli J
28	8	2.4	351	24	Single-stranded DN
29	8	2.4	380	24	Rat sequence diffe
30	8	2.4	403	22	Human immune/haema
31	8	2.4	441	24	DNA encoding human
32	8	2.4	470	22	Human polynucleoti
33	8	2.4	470	22	Wheat tryptophan s
34	8	2.4	475	20	Human immune/haema
35	8	2.4	482	22	Human prostate exp
36	8	2.4	494	23	Human immune/haema
37	8	2.4	509	22	Human immune/haema
38	8	2.4	574	22	Primer specific fo
39	8	2.4	601	22	Primer specific fo
40	8	2.4	639	23	Drosophila melanog
41	8	2.4	683	24	Oligonucleotide fo
42	8	2.4	683	24	Oligonucleotide fo
43	8	2.4	708	22	Human cDNA clone (
44	8	2.4	739	22	Primer specific fo
45	8	2.4	787	22	Human neuroblastom
			859	20	Human secreted pro

ALIGNMENTS

RESULT 1  
AAZ46101  
ID AAZ46101 standard; DNA; 1416 BP.  
XX  
AC AAZ46101;  
XX  
DT 05-MAY-2000 (first entry)  
XX  
DE Nucleotide sequence of the coding region of N1T1 gene.  
XX  
KW N1T1 gene; nitrilase; tumour suppressor gene; FH1T; chromosome 3p14.2;  
XX FRA3B; cancer; genome allele inactivation; ss.  
OS Homo sapiens.  
OS Mus sp.  
OS Drosophila melanogaster.  
XX Caenorhabditis elegans.

```

FH Key      Location/Qualifiers
FT CDS      3..1415
FT          /*tag= a
FT          /product= (pos: 18..20, aa: Xaa)
FT          /product= (pos: 1179..1181, aa: Gln)
FT          /product= (pos: 1182..1184, aa: Ala)
FT          /note= "contains 7 internal stop codons; Xaa is an
FT          unspecified amino acid"
XX
XX WO200003685-A2.
XX
XX 27-JAN-2000.
XX
XX 20-JUL-1999; 99WO-US16366.
XX
XX 20-JUL-1998; 98US-0093350.
XX
XX (UJYE-) UNIV JEFFERSON THOMAS.
XX
XX Croce CM;
XX
XX WPI: 2000-171195/15.
XX P-PSDB: AAY68739.
XX
XX Novel nitrilase homologs used as diagnostic and therapeutic reagents
XX for the detection and treatment of cancer -
XX
XX Claim 6; Fig 6; 25pp; English.
XX
XX The present sequence represents the coding region of human, murine,
XX Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The
XX human and mouse NIT1 genes are members of an uncharacterised
XX mammalian gene family with homology to bacterial and plant nitrilases.
XX The tumour suppressor gene FHIT in D. melanogaster and C. elegans code
XX for fusion proteins in which the Phit domain is fused with a Nit domain.
XX In mouse and humans, FHIT and Nit are encoded by two different genes,
XX localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The
XX human FHIT gene at chromosome 3p14.2, spanning the constitutive
XX chromosomal fragile site FRA3B, is often altered in most common forms
XX of human cancer. The Nit1 protein overcomes the mutated inactivation
XX of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives
XX and analogues of them, and antibodies are used as diagnostic and
XX therapeutic reagents for the detection and treatment of cancers.
XX
XX Sequence 1416 BP; 325 A; 397 C; 363 G; 330 T; 1 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 1416
XX Score: 326.00 Matches: 326
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 99.69% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-357-675c-21 (1-327) x AAZ46101 (1-1416)
XX
XX QY 2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysProGlyLeuArg 21
XX 111 CTGGGCTTCATCACCAGGCTCTCTCACAGATTCCTGTCTCCCTGTGTGCTGGACTCCGG 170
XX
XX QY 22 IleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSerSer 41
XX 171 ATACCTCAACTCTCAGTACTTTGTCTCAGGCCAGGCCAGAGCCATGGCTATCTCCCTCT 230
XX
XX QY 42 SerSerCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLysGln 61
XX 231 TCCTCTCGGAACCTGCCCTGTGTGCTGTGCCAGGTAAACATCGACGCCACAGCAAGCAA 290
XX
XX QY 62 GlnAsnPheLysThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyValaCysLeu 81
XX 291 CAGAACTTTAAACATGTGCTAGCTGTCTCAGAGGCTGCCAGAGCTGGGTGCTGCCTG 350
XX
XX QY 82 AlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeu 101

```

```

Db 351 GCTTTCCTCGCTGAGGCATTTGACTTTCACGGGACCTTGACAGACGCTACACCTG 410
QY 102 SerGluProLeuGlyGlyLeuLeuGluGluTyrThrGlnLeuAlaArgGlyCysGly 121
Db 411 TCTGAACACACCTGGGTGGGAAACTTTTGGGAAGAAATACACCCAGCTGGCCAGGAATGTGGA 470
QY 122 LeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGlnLys 141
Db 471 CTCGTGGCTGCTCTTGGGTGGTTTCCATCAGCGGTGCCCAAGACTGGGACGACACTCAGAAA 530
QY 142 IleTyrAsnCysHisValLeuLeuAsnSerLysGlyAlaValValAlaThrTyrArgLys 161
Db 531 ATCTACAATTTGCTACGTGCTGCTGAACACGAAAGGGGCGAGTAGTGGCCACTTACAGGAAG 590
QY 162 ThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSerThr 181
Db 591 ACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGGCGCTATGTGTGAAGGAACCTCTACC 650
QY 182 MetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeuAla 201
Db 651 ATGCTGGGCCAGCTCTTGAGTCACTGTCAGCACACCCAGCAGCAAGATTGGTCTAGCT 710
QY 202 ValCysTyrAspMetArgPheProGlnLeuSerLeuAlaLeuAlaGlnAlaGlyAlaGlu 221
Db 711 GTCTGCTATGACATGCGGTTCCTCAACTCTCTCTGGCATTTGGCTCAAGCTGGAGCAGAG 770
QY 222 IleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluValLeu 241
Db 771 ATACTTACCTTACCTTTCAGCTTTTGGATCCATTACAGGCCCGCCACTGGGAGGTGTG 830
QY 242 LeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaAlaGlnCysGlyArg 261
Db 831 CTGGGGCCCGTGTATCGAAACCCAGTGCCTATGTAGTGGCAGCAGCACAGCTGTGGAGCC 890
QY 262 HisHisGluLysArgAlaSerTyrGlyHisSerMetValValAlaAspProTrpGlyThrVal 281
Db 891 CACCATGAGAAAGAGAGCAAGTATATGGCCACAGCATGTGTGGTAGACCCCTGGGGAACAGTG 950
QY 282 ValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyrLeu 301
Db 951 GTGGCCCGCTCTCTGAGGGGCCAGGCCCTCTGCCCTTGGCCCAATAGACCTCAACTATCTG 1010
QY 302 ArgGlnLeuArgArgHisLeuProValPheGlnHisArgArgProAspLeuTyrGlyAsn 321
Db 1011 CGACAGTTGGCCGCGACACCTGCCTGTGTTCACGACCGCAGGCCCTGACCTCTATGGCAAT 1070
QY 322 LeuGlyHisProLeuSer 327
Db 1071 CTGGGTACCCCACTGTCT 1088
XX
XX AAF16257 standard; cDNA; 1382 BP.
XX AAF16257;
XX AC AAF16257;
XX XX
XX DF 13-MAR-2001 (first entry)
XX XX
XX DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:692.
XX XX
XX KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
XX KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
XX KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
XX KW antibacterial; gene therapy; neural; immune; reproductive; renal;
XX KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX KW wound; infectious disease; ss.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200055174-A1.
XX XX
XX PD 21-SEP-2000.

```

XX 08-MAR-2000; 2000WO-US05988.  
XX  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX  
XX Rosen CA, Ruben SM;  
PI  
XX WPI; 2000-587513/55.  
DR P-PSDB; AAB57054.  
XX  
PT Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -  
XX  
XX Claim 1; Page 1124; 2339pp; English.  
PS  
XX AAF1566 to AAF1605 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.  
XX  
SQ Sequence 1382 BP; 322 A; 380 C; 357 G; 323 T; 0 other;

## Alignment Scores:

Pred. No.: 1e-247 Length: 1382  
Score: 253.00 Matches: 326  
Percent Similarity: 99.69% Conservative: 0  
Best Local Similarity: 99.69% Mismatches: 1  
Query Match: 77.37% Indels: 1  
DB: 21 Gaps: 0

US-09-357-675C-21 (1-327) x AAF16257 (1-1382)

QY 1 MetLeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysProGlyLeu 20  
DB 77 ATGCTGGGCTTACACAGGCTCTCACAGATTCTGTGCTCTGTGGACTC 136  
QY 21 ArgIleProGlnLeuSerValLeuCysAlaGlnProArgProArgAlaMetAlaIleSer 40  
DB 137 CGGATACCTCAACTCTCAGTACTTTGTGCTACGCCAGGCCAGCCATGGCTATCTCC 196  
QY 41 SerSerSerCysGluLeuProLeuValAlaValCysGlnValThrSerThrProAspLys 60  
DB 197 TCTTCTCTCTGGGAAGTGCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 256  
QY 61 GlnGlnAspPheIleThrCysAlaGluLeuValArgGluAlaAlaArgLeuGlyAlaCys 80  
DB 257 CAACAGAACTTTAAACATGTCTGAGCTGGTTCGAGAGGCTGCCAGACTGGTGCCTGC 316  
QY 81 LeuAlaPheLeuProGluAlaPheAspPheIleAlaArgAspProAlaGluThrLeuHis 100  
DB 317 CTGGCTTTCTGCTGAGGCAATTTGACTTTCATTCAGCGGAGCCCTGCAGAGAGCTACAC 376  
QY 101 LeuSerGluProLeuGlyGlyLysLeuLeuGluGluTyThrGlnLeuAlaArgGluCys 120  
DB 377 CTGCTGTAACCACTGGTGGGAACTTTTGGAAATAATACACCCAGCTTCCAGGGAATCT 436  
QY 121 GlyLeuTrpLeuSerLeuGlyGlyPheHisGluArgGlyGlnAspTrpGluGlnThrGln 140  
DB 437 GGACTCTGGCTTCCTTGGTGGTGTTCATGAGCGGTGGCCAAAGACTGGGAGCAGACTCAG 496

QY 141 LysIleTyrAsnCysHisValLeuLeuAsnSerLysGlyValaValValAlaThrTyrArg 160  
DB 497 AAATCTACATTTGTCACGTGCTGTAACAGCAAGAGGGCAGTAGTGCCACTTACAGG 556  
QY 161 LysThrHisLeuCysAspValGluIleProGlyGlnGlyProMetCysGluSerAsnSer 180  
DB 557 AAGACACATCTGTGTGACGTAGAGATTCCAGGGCAGGGCT-ATGTGTGAAAGCAACTCT 615  
QY 181 ThrMetProGlyProSerLeuGluSerProValSerThrProAlaGlyLysIleGlyLeu 200  
DB 616 ACCATGCTGGGCCAGCTTGTGATCCTGTACGACACACAGCAAGATTTGGTCTA 675  
QY 201 AlaValCysTyrAspMetArgPheProGluLeuSerLeuAlaLeuAlaGlnAlaGlyAla 220  
DB 676 GCTGTCTGTATGACATGCGGTTCCTGAACCTCTCTCTGGCATTTGCTCAAGCTTGAGCA 735  
QY 221 GluIleLeuThrTyrProSerAlaPheGlySerIleThrGlyProAlaHisTrpGluVal 240  
DB 736 GAGATACCTTACCTATCCTTTTGGATCCATTACAGGCCAGCCACTGGGAGGTG 795  
QY 241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrValValAlaAlaGlnCysGly 260  
DB 796 TTGCTGCGGGCCGCTGTATCGAAACCCAGTGTATGTAGTGGCAGCAGCAGCTGTGA 855  
QY 261 ArgHisHisGlyLysArgAlaSerTyrGlyHisSerMetValValAspProTrpGlyThr 280  
DB 856 CGCCACCATGAGAGAGACCAAGTTATGCGCCACAGCATGGTGTAGACCCCTGGGGAACA 915  
QY 281 ValValAlaArgCysSerGluGlyProGlyLeuCysLeuAlaArgIleAspLeuAsnTyr 300  
DB 916 GTGGTGGCCGCTGCTCTGAGGGGCCAGGCCTCTGCTTGCCTGCGGAATAGACCTCAACTAT 975  
QY 301 LeuArgGlnLeuArgHisLeuProValPheGlnHisArgProAspLeuTyrGly 320  
DB 976 CTGCGACAGTTGCGCGACACCTGCTGTGTTCCAGACCCGAGGCTTACCTCTATGGC 1035  
QY 321 AsnLeuGlyHisProLeuSer 327  
DB 1036 AATCTGGGTCAACCACTGTCT 1056  
RESULT 3  
AAAX30398  
ID AAAX30398 standard; DNA; 1203 BP.  
XX  
AC AAAX30398;  
XX  
DT 14-MAY-1999 (first entry)  
XX  
DE DNA encoding a human secreted protein.  
XX  
KW Secreted protein; cancer; tumour; neurodegenerative disorder;  
KW developmental abnormality; foetal deficiency; blood disorder;  
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;  
KW renal disease; diabetes; inflammation; allergy; ischemic shock;  
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;  
KW prostate disease; asthma; osteoporosis; arthritis; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO9907891-Al.  
XX  
XX 18-FEB-1999.  
XX  
XX 04-AUG-1998; 98WO-US16235.  
XX  
XX 19-AUG-1997; 97US-0056732.  
PR 05-AUG-1997; 97US-0054798.  
PR 05-AUG-1997; 97US-0054803.  
PR 05-AUG-1997; 97US-0054804.  
PR 05-AUG-1997; 97US-0054806.  
PR 05-AUG-1997; 97US-0054807.  
PR 05-AUG-1997; 97US-0054808.



CC The invention describes an isolated polynucleotide and polypeptide  
CC useful for stimulating and/or expanding T cells specific for a tumour  
CC protein for determining the presence of a cancer in a patient. A  
CC composition containing the polynucleotide and/or polypeptide is useful  
CC for treating a lung cancer in a patient. The polypeptide is useful for  
CC removing tumour cells from a biological sample. The polynucleotide is  
CC also useful as probe or primer to detect the level of mRNA encoding a  
CC tumour protein. This sequence encodes a lung tumour associated protein  
CC or protein fragment, described in the method of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 377 BP; 94 A; 100 C; 112 G; 71 T; 0 other;

## Alignment Scores:

Pred. No.: 1.08e-54 Length: 377  
Score: 63.00 Matches: 124  
Percent Similarity: 98.41% Conservative: 0  
Best Local Similarity: 98.41% Mismatches: 1  
Query Match: 19.27% Indels: 2  
DB: 24 Gaps: 0

US-09-357-675C-21 (1-327) x ABK39058 (1-377)

QY 8 ProProHisArgPheLeuSerLeuLeuCysProGlyLeuArgIleProGlnLeuSerVal 27  
Db CCTCTCAGATTCTGTCCTTCTGTCTCTGGACCTCGGATACCTCAACTCTCAGTA 318  
QY 28 LeuCysAlaGlnProArgProArgAlaMetAlaIleSerSerSerCysGluLeuPro 47  
Db CTTTGTGCTCAGCCAGGCCAGAGCATGGCTATCTCTCTCTCTCTCTCTCTCTCTCT 258  
QY 48 LeuValAlaValCysGlnValThrSerThrProAspLysGlnAsnPhelysThrCys 67  
Db CTGGTGGCTGTGTCAGGTATACGTCGCGCCAGACAGCAACAGCAACATGTT 198  
QY 68 AlaGlu-LeuValArgGluAlaAlaArgLeuGlyAlaCysLeuAlaPheLeuProGluAl 87  
Db GCTGG-GCTGGTTCGAGAGCTGCCAGACTGGTGGCTGCTGCTGCTGCTGCTGCTGCTG 139  
QY 87 aPheAspPheIleAlaArgAspProAlaGluThrLeuHisLeuSerGluProLeuGlyGl 107  
Db ATTGACTTCATTGACGGGACCCCTGCAGACGCTACACCTGTCTGAACCACTGGGTGG 79  
QY 107 YLysLeuLeuGluGluTyThrGlnLeuAlaArgGluCysGlyLeuTrpLeuSerLeuGl 127  
Db GAAACTTTTGGGAAGATAACACCCAGCTTGCCAGGGAATGTGCACTGTGCTCTCTCTGG 19  
QY 127 YGlyPheHisGluArg 132  
Db TGGTTTCATGAGCGT 3  
RESULT 5  
ID AAD25458  
XX AAD25458 standard; cDNA; 1214 BP.  
AC AAD25458;  
XX  
DT 26-MAR-2002 (first entry)  
DE  
DE Xenopus laevis Nit1 cDNA.  
XX  
XX Frog; Nit1; cytostatic; neuroprotective; cellular pathway; therapy;  
KW apoptosis; proliferative disorder; degenerative disease; ss.  
XX  
OS Xenopus laevis.  
XX  
FH Key Location/Qualifiers  
FT CDS 180...1046  
FT /\*tag= a  
FT /product= "Frog Nit1 protein"

XX WO200187958-A2.  
PN 22-NOV-2001.  
PD 15-MAY-2001; 2001WO-US15664.  
PF 16-MAY-2000; 2000US-204713P.  
PR (UYJE-) UNIV JEFFERSON THOMAS.  
PA Croce C, Brenner C, Pekarski Y;  
PI WPI; 2002-082984/11.  
DR P-PSDB; AAE15792.  
XX Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins, useful to find molecules that mimic or antagonize Fhit interaction for the treatment of proliferative or degenerative diseases  
PT Claim 21; Page 60; 61pp; English.  
CC The invention relates to isolated nucleic acids comprising a fully defined cDNA nucleotide sequence encoding human, Xenopus laevis and mouse Nit2 proteins. Nit and Fhit proteins are encoded as fusion proteins in invertebrates and as separate polypeptides in vertebrates. Nit and Fhit interact physically and as separate polypeptides in same cellular pathways. Molecules which bind Nit2 and mimic or antagonise Fhit interaction are used to treat diseases in which activity of Nit2 protein is altered in a mammal. Fhit mimics induce apoptosis and are particularly useful to treat proliferative disorders, whilst Fhit antagonists promote cell proliferation and are particularly useful to treat degenerative disease. The present sequence is frog Nit1 cDNA.  
SQ Sequence 1214 BP; 281 A; 305 C; 345 G; 282 T; 1 other;

## Alignment Scores:

Pred. No.: 1.93e-13 Length: 1214  
Score: 23.00 Matches: 23  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.03% Indels: 0  
DB: 24 Gaps: 0

US-09-357-675C-21 (1-327) x AAD25458 (1-1214)

QY 236 AlaHisTrpGluValLeuLeuArgAlaAlaIleGluThrGlnCysTyrValValAla 255  
Db 765 GCACATTGGAGGTGTGCTGAGAGCCCGTGCATAGAAACCCAGTGTACGTAGTTGCA 824

QY 256 AlaAlaGln 258

Db 825 GCGGCACAG 833

## RESULT 6

AAK61236  
ID AAK61236 standard; cDNA; 539 BP.  
XX  
AC AAK61236;  
XX

DT 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6296.

DE  
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ss.  
XX  
OS Homo sapiens.  
XX  
FH WO200157182-A2.  
FT CDS  
FT 09-AUG-2001.

[illegible]

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
PS Claim 1; SEQ ID NO 6296; 307lpp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 539 BP; 129 A; 117 C; 137 G; 152 T; 4 other;

Alignment Scores:  
Pred. No.: 1.08e-08 Length: 539  
Score: 18.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.50% Indels: 0  
DB: 22 Gaps: 0

US-09-357-675C-21 (1-327) x AAK61236 (1-539)

QY 2 LeuGlyPheIleThrArgProHisArgPheLeuSerLeuLeuCysProGly 19  
|||||  
Db 373 CTGGGCTTCACCCAGGCGCTCTCACAGATTCCTGCTCTGTGTCCKGKN 426

RESULT 7  
ABQ66352  
ID ABQ66352 standard; DNA; 604 BP.  
XX  
AC ABQ66352;  
XX  
DT 21-AUG-2002 (first entry)  
XX  
DE Arabidopsis thaliana polynucleotide SEQ ID NO 929.  
XX  
KW Arabidopsis thaliana; thale cress; plant; transgenic; GMO; disease;  
KW stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;  
KW insecticide; antibiotic; ds.  
XX  
OS Arabidopsis thaliana.

XX  
XX US2002059663-A1.  
XX  
PN 16-MAY-2002.  
PD  
XX  
PF 26-JAN-2001; 2001US-0770149.  
XX  
PR 27-JAN-2000; 2000US-178506P.  
XX  
XX (GORLACH J.  
PA (ANY) AN Y.  
PA (HAMIL) HAMILTON C M.  
PA (PRIC) PRICE J L.  
PA (RAIN) RAINES T M.  
PA (YUY) YU Y.  
PA (RAME) RAMEAKA J G.  
PA (PAGE) PAGE A.  
PA (MATH) MATHAW A V.

PA (LEDF) LEDFORD B L.  
PA (WOES) WOESSNER J P.  
PA (HAAS) HAAS W D.  
PA (GARC) GARCIA C A.  
PA (KRIC) KRICKER M.  
PA (SLAT) SLATER T.  
PA (DAVI) DAVIS K R.  
PA (ALLE) ALLEN K.  
PA (HOFF) HOFFMAN N.  
PA (HURB) HURBAN P.  
XX  
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;  
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;  
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;  
PI Hurban P;  
XX WPI; 2002-479224/51.  
XX  
PT New nucleic acid that hybridizes to Arabidopsis thaliana sequences,  
PT useful e.g. for preparing transgenic plants with increased resistance  
PT or altered metabolism -  
XX  
PS Claim 1; SEQ ID NO 929; 40pp + Sequence Listing; English.  
XX  
CC The invention relates to nucleic acids (I) that hybridise under stringent  
CC conditions to any of 999 sequences (ABQ65424-ABQ66422) or their  
CC fragments. (I) are used to express the corresponding polypeptides (II) or  
CC to produce genetically modified plant cells or transgenic plants, which  
CC may have improved resistance to disease or stress, or altered  
CC metabolic/biosynthetic pathways (for production of commercial,  
CC nutritional or medicinal products), or generally any trait of interest,  
CC or can be used to screen for biologically active agents (e.g. fungicides,  
CC insecticides and antibiotics).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO at seqdata.uspto.gov/sequence.html?docid=999909770149.  
XX  
SQ Sequence 604 BP; 174 A; 113 C; 139 G; 176 T; 2 other;

Alignment Scores:  
Pred. No.: 0.0015 Length: 604  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.98% Indels: 0  
DB: 24 Gaps: 0

US-09-357-675C-21 (1-327) x ABQ66352 (1-604)

QY 241 LeuLeuArgAlaArgAlaIleGluThrGlnCysTyrVal 253  
|||||  
Db 198 CTCTTCGAGCCCGAGCAATTGAACTCAATGTTATGTC 236

RESULT 8  
ABL12225  
ID ABL12225 standard; cDNA; 1495 BP.  
XX  
AC ABL12225;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 31157.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX P-PSDB; ABB68122.  
DR New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 31157; 2lpp + Sequence Listing; English.  
PS The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 1495 BP; 388 A; 381 C; 391 G; 335 T; 0 other;  
Alignment Scores:  
Pred. No.: 0.398 Length: 1495  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.36% Indels: 0  
DB: 23 Gaps: 0  
US-09-357-675C-21 (1-327) x ABL12225 (1-1495)  
QY 241 LeuLeuArgAlaArgAlaIleGluThrGlnCys 251  
DB 704 CTCCTGCGGGCCAGAGCCATAGAGACTCAATGC 736  
RESULT 9  
ABL12224  
ID ABL12224 standard; cDNA; 3548 BP.  
XX ABL12224;  
AC ABL12224;  
XX 26-MAR-2002 (first entry)  
DT Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 31154.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
XX 27-SEP-2001.  
PD 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 12139; 2lpp + Sequence Listing; English.

DR WPI; 2001-656860/75.  
DR P-PSDB; ABB68121.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 31154; 2lpp + Sequence Listing; English.  
PS The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 3548 BP; 952 A; 842 C; 864 G; 890 T; 0 other;  
Alignment Scores:  
Pred. No.: 0.93 Length: 3548  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.36% Indels: 0  
DB: 23 Gaps: 0  
US-09-357-675C-21 (1-327) x ABL12224 (1-3548)  
QY 241 LeuLeuArgAlaArgAlaIleGluThrGlnCys 251  
DB 1757 CTCCTGCGGGCCAGAGCCATAGAGACTCAATGC 1789  
RESULT 10  
ABL20222  
ID ABL20222 standard; DNA; 5692 BP.  
XX ABL20222;  
AC ABL20222;  
XX 26-MAR-2002 (first entry)  
DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12139.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
XX 27-SEP-2001.  
PD 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX Claim 1; SEQ ID NO 12139; 2lpp + Sequence Listing; English.



CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (AB057737-AB072072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 5692 BP; 1509 A; 1354 C; 1405 G; 1424 T; 0 other;

Alignment Scores:  
Pred. No.: 1.48 Length: 5692  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.36% Indels: 0  
DB: 23 Gaps: 0

US-09-357-675C-21 (1-327) x ABL20222 (1-5692)

QY 241 LeuLeuArgAlaArgAlaIleGluThrGlnCys 251  
|||||  
DB 3416 CTCCTCGGGCCAGAGCCATAGAGACTCAATGC 3448

RESULT 11  
ABN23783  
ID ABN23783 standard; cDNA; 242 BP.  
XX  
AC ABN23783;

24-JUN-2002 (first entry)

DE Human ORFX polynucleotide sequence SEQ ID NO:15043.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis; gene; ss.

XX Homo sapiens.

PN WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US10836.

XX 30-MAY-2000; 2000US-206132P.

XX 29-AUG-2000; 2000US-228716P.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach MD;

XX WPI; 2002-106308/14.

DR P-PSDB; ABP08031.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders

XX Disclosure; SEQ ID 16043; 1037pp; English.

XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP1500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic  
CC transplantation, cardiovascular diseases, disorders related to organ  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 242 BP; 78 A; 48 C; 52 G; 63 T; 1 other;

Alignment Scores:  
Pred. No.: 7.26 Length: 242  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.75% Indels: 0  
DB: 24 Gaps: 0

US-09-357-675C-21 (1-327) x ABN23783 (1-242)

QY 203 CysTyrAspMetArgPheProGluLeu 211

|||||  
DB 199 TGCTATGACATGAGGTTTCGGAACTC 225

RESULT 12

AAH19580/c

ID AAH19580 standard; cDNA; 6597 BP.

XX AAH19580;

XX 26-JUL-2001 (first entry)

DE Human polynucleotide #3 expressed in intraabdominal adipose tissue.

XX Human; intraabdominal adipose tissue; gene expression; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 95..6367

FT /\*tag= a

XX JP2001008699-A.

XX 16-JAN-2001.

XX 30-JUN-1999; 99JP-0185737.

XX 30-JUN-1999; 99JP-0185737.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX WPI; 2001-275911/29.

XX P-PSDB; AAB97070.

XX Assuming the intraabdominal adipose tissue amount, comprising relating  
PT the expression level of a gene expression product to the amount -

XX  
PS Claim 1; Page 15-23; 33pp; Japanese.  
CC The present sequence is provided in a specification relating to a  
CC method for predicting the amount of intraabdominal adipose tissue.  
CC The method involves relating the amount of adipose tissue to the  
CC expression level of at least one gene transcription product which  
CC has a 491 or 2090 residue amino acid sequence and is encoded by a  
CC 2376 or 2090 base pair sequence, or which can hybridize to a 2385  
CC or the 2090 base pair sequence under stringent conditions. The  
CC method includes a step of deriving the area value of the  
CC intraabdominal adipose tissue at the cross section of abdominal  
CC navel, from the expression level of a gene transcription product  
CC in a sample of the intraabdominal adipose tissue. The method can  
CC be used for predicting the amount of intraabdominal adipose  
CC tissue irrespective of the amount of subcutaneous adipose tissue.  
CC The present sequence encodes a gene transcription product whose  
CC expression level may be measured as part of this method.  
XX  
SQ Sequence 6597 BP; 1590 A; 1870 C; 1549 G; 1588 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 186 Length: 6597  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.75% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-357-675C-21 (1-327) x AAH19580 (1-6597)  
  
Qy 212 SerLeuAlaLeuAlaGlnAlaGlyAla 220  
|||||  
Db 2660 TCTCTCGCTCTGGCACAGCAGTGCC 2634  
|||||  
  
RESULT 13  
ID ABK84618/c  
ID ABK84618 standard; cDNA: 6597 BP.  
XX  
AC ABK84618;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Human cDNA differentially expressed in granulocytic cells #1189.  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US30821.  
XX  
PR 03-OCT-2000; 2000US-237189P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX  
DR WPI; 2002-435328/46.  
XX  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and

PT drug toxicity  
XX  
PS Claim 1; SEQ ID No 1189; 114pp; English.  
XX  
CC The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6597 BP; 1590 A; 1870 C; 1549 G; 1588 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 186 Length: 6597  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.75% Indels: 0  
DB: 24 Gaps: 0  
  
US-09-357-675C-21 (1-327) x ABK84618 (1-6597)  
  
Qy 212 SerLeuAlaLeuAlaGlnAlaGlyAla 220  
|||||  
Db 2660 TCTCTCGCTCTGGCACAGCAGTGCC 2634  
|||||  
  
RESULT 14  
AAA81479/c  
ID AAA81479 standard; DNA: 69936 BP.  
XX  
AC AAA81479;  
XX  
DT 04-DEC-2000 (first entry)  
XX  
DE N. meningitidis partial DNA sequence gnm\_27 SEQ ID NO:27.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.  
XX  
OS Neisseria meningitidis.  
XX

```

PN WO200022430-A2.
XX
XX 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US23573.
XX
XX 09-OCT-1998; 98US-0103794.
XX
XX 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX Rappuoli R, Pizza M;
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
XX used in the diagnosis and treatment of N. meningitidis infection and
XX other Neisserial infections, for example, N.gonorrhoea
XX
XX Claim 7; Page 547-567; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies
XX against them, can be used in the manufacture of a composition. The
XX composition can be used as a medicament (or in the manufacture of a
XX medicament) for treating, preventing or diagnosing infection due to
XX Neisserial bacteria. For example, some of the identified proteins could
XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neisseriae. Identification of sequences
XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious
XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX Multivalent vaccines have also been tried but none have successfully
XX overcome antigenic variability. The provision of further, complete
XX sequences may provide an opportunity to identify secreted or surface
XX exposed proteins that may be presumed targets for the immune system and
XX which are not antigenically variable or at least more conserved than
XX other more variable regions.
XX
XX Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.89e+03 Length: 69936
XX Score: 9.00 Matches: 9
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 2.75% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-357-675C-21 (1-327) x AAA81479 (1-69936)
XX
XX QY 250 GlnCysTyrValValAlaAlaGln 258
XX |||||
XX DB 26543 CAATGTAGTCGTGGCGGCACAG 26517
XX
XX RESULT 15
XX AAF21607
XX ID AAF21607 standard; DNA; 349980 BP.
XX
XX AC AAF21607;
XX
XX DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX
XX Neisseria meningitidis; Neisseria gonorrhoea; immunogenic; vaccine;
XX diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX ds.
XX
XX Neisseria meningitidis.
XX
XX WO2000066791-A1.
XX
XX 09-NOV-2000.
XX
XX 08-MAR-2000; 2000WO-US05928.
XX
XX 30-APR-1999; 99US-0132068.
XX
XX 08-OCT-1999; 99WO-US23573.
XX
XX 28-FEB-2000; 2000GB-0004695.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
XX Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
XX Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
XX frames are used to detect, treat and prevent Neisserial infections -
XX Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of
XX Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
XX to AAF21613 represent fragments of the NMB genomic sequence, as the
XX sequence was too long to go in a record on its own it was split into 8
XX sequences which overlap each other at the beginning and end of each
XX sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
XX the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
XX the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
XX Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
XX AAF21606 represent PCR primers which are used in the exemplification of
XX the present invention. The NMB genome and fragments from it have
XX antibacterial activity, and can be used in vaccines and gene therapy.
XX Neisseria nucleic acids, proteins and/or antibodies which binds to the
XX proteins can be used in compositions for treating or preventing infection
XX due to Neisserial bacteria or as a diagnostic reagent for detecting the
XX presence of Neisserial bacteria or of antibodies raised to Neisserial
XX bacteria. Computers, computer memory, computer storage medium or computer
XX databases can be used in a search to identify open reading frames (ORFs)
XX or coding sequences within the NMB genome. The DNA sequences provide
XX further opportunities to find antigenic or immunogenic proteins which are
XX more effective in vaccines than the outer membrane proteins currently
XX used.
XX
XX Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 9.2e+03 Length: 349980
XX Score: 9.00 Matches: 9
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 2.75% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-357-675C-21 (1-327) x AAF21607 (1-349980)
XX
XX QY 250 GlnCysTyrValValAlaAlaGln 258
XX |||||
XX DB 154644 CAATGTAGTCGTGGCGGCACAG 154670
XX
XX Search completed: December 8, 2002, 10:45:31

```

Job time : 368 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 8, 2002, 10:38:31 ; Search time 52 seconds

(without alignments)  
1928.525 Million cell updates/sec

Title: US-09-357-675C-21

Perfect score: 327

Sequence: 1 MLGFTTPPHRFLSLCPGL.....LPVFQHRPDLVGNLGHPLS 327

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878600

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n\_model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09357675/runat\_03122002\_142236\_20423/app\_query.fasta\_1.519  
-DB=IssuedPatents\_NA -QEXT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09357675 -CGCN\_1\_1\_17 @runat\_03122002\_142236\_20423 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGESQUEY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMBOUR=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAEXT=60 -FGAPOP=6 -FGAEXT=7  
-YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/l/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/l/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/l/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/l/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/l/ina/PCTUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/l/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.4	56	1	US-08-173-968-6
2	8	2.4	56	3	Sequence 6, Appli
3	8	2.4	115	4	Sequence 6, Appli
4	8	2.4	351	4	Sequence 18, Appl
5	8	2.4	864	4	Sequence 17, Appl
6	8	2.4	864	4	Sequence 12, Appl
7	8	2.4	918	4	Sequence 13, Appl
8	8	2.4	918	4	Sequence 5, Appli
9	8	2.4	960	4	Sequence 5, Appli
10	8	2.4	960	4	Sequence 36, Appl
11	8	2.4	990	4	Sequence 35, Appl
12	8	2.4	990	4	Sequence 34, Appl

c 13	8	2.4	1335	4	US-09-439-261-1	Sequence 1, Appli
c 14	8	2.4	1335	4	US-09-227-613-1	Sequence 1, Appli
c 15	8	2.4	1508	4	US-09-039-046-1	Sequence 1, Appli
c 16	8	2.4	1881	4	US-08-553-498-1	Sequence 1, Appli
c 17	8	2.4	1881	4	US-09-553-498-3	Sequence 3, Appli
c 18	8	2.4	1881	4	US-09-618-869-1	Sequence 1, Appli
c 19	8	2.4	1881	4	US-09-618-869-3	Sequence 3, Appli
c 20	8	2.4	1946	4	US-08-755-584-1	Sequence 1, Appli
c 21	8	2.4	1946	3	US-09-192-611-1	Sequence 1, Appli
c 22	8	2.4	2159	4	US-09-039-046-4	Sequence 4, Appli
c 23	8	2.4	3300	1	US-08-194-230-6	Sequence 6, Appli
c 24	8	2.4	3300	2	US-08-614-377A-6	Sequence 6, Appli
c 25	8	2.4	3300	4	US-09-142-648B-6	Sequence 6, Appli
c 26	8	2.4	3350	1	US-08-247-946A-2	Sequence 2, Appli
c 27	8	2.4	3350	5	PCT-US95-06420-2	Sequence 2, Appli
c 28	8	2.4	4085	1	US-08-072-574-7	Sequence 7, Appli
c 29	8	2.4	4085	1	US-08-486-270-7	Sequence 7, Appli
c 30	8	2.4	4085	3	US-08-367-264-7	Sequence 7, Appli
c 31	8	2.4	4085	4	US-09-153-757-7	Sequence 7, Appli
c 32	8	2.4	4181	1	US-08-072-574-9	Sequence 9, Appli
c 33	8	2.4	4181	1	US-08-486-270-9	Sequence 9, Appli
c 34	8	2.4	4181	3	US-08-367-264-9	Sequence 9, Appli
c 35	8	2.4	4181	4	US-09-153-757-9	Sequence 9, Appli
c 36	8	2.4	4207	4	US-08-660-148-1	Sequence 1, Appli
c 37	8	2.4	4207	4	US-08-660-148-3	Sequence 3, Appli
c 38	8	2.4	4303	4	US-08-660-148-4	Sequence 4, Appli
c 39	8	2.4	4303	4	US-08-660-148-6	Sequence 6, Appli
c 40	8	2.4	4326	2	US-08-852-807-12	Sequence 12, Appl
c 41	8	2.4	4817	1	US-07-951-715A-18	Sequence 18, Appl
c 42	8	2.4	4817	2	US-08-459-448A-18	Sequence 18, Appl
c 43	8	2.4	4817	3	US-08-459-595A-18	Sequence 18, Appl
c 44	8	2.4	4817	3	US-08-459-504B-18	Sequence 18, Appl
c 45	8	2.4	4817	3	US-08-459-444-18	GENERAL INFORMA

ALIGNMENTS

RESULT 1

US-08-173-968-6

; Sequence 6, Application US/08173968

; Patent No. 5576292

; GENERAL INFORMATION:

; APPLICANT: Weiss, Jerrold

; TITLE OF INVENTION: Biologically Active

; TITLE OF INVENTION: Bactericidal/Permeability-increasing Protein Fragments

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howard M. Frankfort

; STREET: 805 Third Ave.

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/173,968

; FILING DATE:

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/754,204

; FILING DATE: 26-AUG-1991

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Frankfort, Howard M.

; REGISTRATION NUMBER: 32,613

; REFERENCE/DOCKET NUMBER: 5969/05982US1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 527-7700

```
;
; TELEFAX: (212) 753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..56
; OTHER INFORMATION: notes-This sequence can be found
; OTHER INFORMATION: in the application, as filed on page 25, line 28.
; PUBLICATION INFORMATION:
; AUTHORS: Ooi, C. E. et al.
; JOURNAL: J. Biol. Chem.
; VOLUME: 262
; PAGES: 14891-14894
; DATE: 1987
; US-08-173-968-6
;
; Alignment Scores:
; Pred. No.: 5.49 Length: 56
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 2.45% Indels: 0
; DB: 1 Gaps: 0
;
; US-09-357-675C-21 (1-327) x US-08-173-968-6 (1-56)
;
; QY 286 SerGluGlyProGlyLeuCysLeu 293
; Db 30 TCAGAAGGGCGCTGGATTATGCTC 53
;
; RESULT 2
; US-07-801-814D-6
; Sequence 6, Application US/07801814D
; Patent No. 6132775
; GENERAL INFORMATION:
; APPLICANT: Elsbach, Peter
; APPLICANT: Weiss, Jerrold
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE
; TITLE OF INVENTION: BACTERICIDAL/PERMEABILITY-INCREASING
; TITLE OF INVENTION: PROTEIN FRAGMENTS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby P.C.
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch,
; MEDIUM TYPE: 360 Kb storage
; COMPUTER: IBM or IBM-compatible
; OPERATING SYSTEM: PC/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/801,814D
; FILING DATE: 19911204
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Emily Miao
; REGISTRATION NUMBER: 35,285
; REFERENCE/DOCKET NUMBER: 5986/13611-US3
```

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: This sequence can be
; OTHER INFORMATION: found in the application, as filed, on
; OTHER INFORMATION: page 23, line 6.
; PUBLICATION INFORMATION:
; AUTHORS: Ooi, C. E. et al.
; TITLE:
; JOURNAL: J. Biol. Chem.
; VOLUME: 262
; ISSUE:
; PAGES: 14891-14894
; DATE: 1987
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: 1-56
; US-07-801-814D-6
;
; Alignment Scores:
; Pred. No.: 5.49 Length: 56
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 2.45% Indels: 0
; DB: 3 Gaps: 0
;
; US-09-357-675C-21 (1-327) x US-07-801-814D-6 (1-56)
;
; QY 286 SerGluGlyProGlyLeuCysLeu 293
; Db 30 TCAGAAGGGCGCTGGATTATGCTC 53
;
; RESULT 3
; US-09-660-877-18/c
; Sequence 18, Application US/09660877
; Patent No. 6251639
; GENERAL INFORMATION:
; APPLICANT: Kurn, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LINEAR
; TITLE OF INVENTION: ISOTHERMAL AMPLIFICATION OF POLYNUCLEOTIDE SEQUENCES
; FILE REFERENCE: 492692000100
; CURRENT APPLICATION NUMBER: US/09/660,877
; CURRENT FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 115
; TYPE: DNA
; ORGANISM: Synthetic primer
; FEATURE:
; OTHER INFORMATION: IA013
; US-09-660-877-18
;
; Alignment Scores:
; Pred. No.: 10.8 Length: 115
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
```

Query Match: 2.45% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-357-675C-21 (1-327) x US-09-660-877-18 (1-115)  
 QY 189 SerProValSerThrProAlaGly 196  
 DB 105 TCTCCAGTGTCCACCCCTGCCGGG 82  
 RESULT 4  
 US-09-660-877-17/c  
 ; Sequence 17, Application US/09660877  
 ; Patent No. 6251639  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kuro, N.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LINEAR  
 ; FILE REFERENCE: 49269200100  
 ; CURRENT APPLICATION NUMBER: US/09/660,877  
 ; CURRENT FILING DATE: 2000-09-13  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 17  
 ; LENGTH: 351  
 ; TYPE: DNA  
 ; ORGANISM: Synthetic primer  
 US-09-660-877-17  
 Alignment Scores:  
 Pred. No.: 30.8 Length: 351  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.45% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-357-675C-21 (1-327) x US-09-660-877-17 (1-351)  
 QY 189 SerProValSerThrProAlaGly 196  
 DB 105 TCTCCAGTGTCCACCCCTGCCGGG 82  
 RESULT 5  
 US-09-439-261-12/c  
 ; Sequence 12, Application US/09439261  
 ; Patent No. 6428990  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abbott Laboratories  
 ; APPLICANT: Mukerji, Pradip  
 ; APPLICANT: Leonard, Amanda E.  
 ; APPLICANT: Huang, Yung-Sheng  
 ; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF  
 ; FILE REFERENCE: 6295 US.P2  
 ; CURRENT APPLICATION NUMBER: US/09/439,261  
 ; CURRENT FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: US 08/833,610  
 ; PRIOR FILING DATE: 1997-04-11  
 ; PRIOR APPLICATION NUMBER: PCT/US98/07422  
 ; PRIOR FILING DATE: 1998-04-10  
 ; PRIOR APPLICATION NUMBER: US 09/227,613  
 ; PRIOR FILING DATE: 1999-01-08  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 864  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-439-261-12  
 Alignment Scores:  
 Pred. No.: 71.9 Length: 864  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.45% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-357-675C-21 (1-327) x US-09-439-261-12 (1-864)  
 QY 72 ArgGluAlaAlaArgLeuGlyAla 79  
 DB 361 AGAGAGGCAGCAAGGCTGGGGCC 338  
 RESULT 6  
 US-09-227-613-13/c  
 ; Sequence 13, Application US/09227613A  
 ; Patent No. 6432684  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MUKERJI, Pradip  
 ; APPLICANT: LEONARD, Amanda E.  
 ; APPLICANT: HUANG, Yung-Sheng  
 ; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF  
 ; FILE REFERENCE: 6295.US.P1  
 ; CURRENT APPLICATION NUMBER: US/09/227,613A  
 ; CURRENT FILING DATE: 1999-01-08  
 ; PRIOR APPLICATION NUMBER: 08/833,610  
 ; PRIOR FILING DATE: 1997-04-11  
 ; NUMBER OF SEQ ID NOS: 42  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 13  
 ; LENGTH: 864  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapien  
 US-09-227-613-13  
 Alignment Scores:  
 Pred. No.: 71.9 Length: 864  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.45% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-357-675C-21 (1-327) x US-09-227-613-13 (1-864)  
 QY 72 ArgGluAlaAlaArgLeuGlyAla 79  
 DB 361 AGAGAGGCAGCAAGGCTGGGGCC 338  
 RESULT 7  
 US-09-439-261-5/c  
 ; Sequence 5, Application US/09439261  
 ; Patent No. 6428990  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abbott Laboratories  
 ; APPLICANT: Mukerji, Pradip  
 ; APPLICANT: Leonard, Amanda E.  
 ; APPLICANT: Huang, Yung-Sheng  
 ; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF  
 ; FILE REFERENCE: 6295 US.P2  
 ; CURRENT APPLICATION NUMBER: US/09/439,261  
 ; CURRENT FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: US 08/833,610  
 ; PRIOR FILING DATE: 1997-04-11  
 ; PRIOR APPLICATION NUMBER: PCT/US98/07422  
 ; PRIOR FILING DATE: 1998-04-10  
 ; PRIOR APPLICATION NUMBER: US 09/227,613  
 ; PRIOR FILING DATE: 1999-01-08  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 5  
 ; LENGTH: 918  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature

```
; LOCATION: (755)...(755)
; OTHER INFORMATION: r = g or a at position 755
US-09-439-261-5

Alignment Scores:
Pred. No.: 76.1 Length: 918
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-439-261-5 (1-918)

QY 72 ArgGluAlaAlaArgLeuGlyAla 79
|||||
Db 796 AGAGAGCGCAGCAAGCTGGGGGCC 773

RESULT 8
US-09-227-613-5/c
; Sequence 5, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
; APPLICANT: MUKERJI, Pradip
; APPLICANT: LEONARD, Amanda E.
; APPLICANT: HUANG, Yung-Sheng
; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF
; FILE REFERENCE: 6295.US.P1
; CURRENT APPLICATION NUMBER: US/09/227,613A
; CURRENT FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 08/833,610
; PRIOR FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-227-613-5

Alignment Scores:
Pred. No.: 76.1 Length: 918
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-227-613-5 (1-918)

QY 72 ArgGluAlaAlaArgLeuGlyAla 79
|||||
Db 796 AGAGAGCGCAGCAAGCTGGGGGCC 773

RESULT 9
US-09-439-261-36/c
; Sequence 36, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda E.
; APPLICANT: Huang, Yung-Sheng
; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF
; FILE REFERENCE: 6295.US.P2
; CURRENT APPLICATION NUMBER: US/09/439,261
; CURRENT FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 08/833,610
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: PCT/US98/07422
; PRIOR FILING DATE: 1998-04-10
; PRIOR APPLICATION NUMBER: US 09/227,613
; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-439-261-36

Alignment Scores:
Pred. No.: 79.4 Length: 960
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-439-261-36 (1-960)

QY 72 ArgGluAlaAlaArgLeuGlyAla 79
|||||
Db 862 AGAGAGCGCAGCAAGCTGGGGGCC 839

RESULT 10
US-09-227-613-35/c
; Sequence 35, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
; APPLICANT: MUKERJI, Pradip
; APPLICANT: LEONARD, Amanda E.
; APPLICANT: HUANG, Yung-Sheng
; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF
; FILE REFERENCE: 6295.US.P1
; CURRENT APPLICATION NUMBER: US/09/227,613A
; CURRENT FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 08/833,610
; PRIOR FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 35
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-227-613-35

Alignment Scores:
Pred. No.: 79.4 Length: 960
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-227-613-35 (1-960)

QY 72 ArgGluAlaAlaArgLeuGlyAla 79
|||||
Db 862 AGAGAGCGCAGCAAGCTGGGGGCC 839

RESULT 11
US-09-439-261-35/c
; Sequence 35, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda E.
; APPLICANT: Huang, Yung-Sheng
; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF
; FILE REFERENCE: 6295.US.P2
; CURRENT APPLICATION NUMBER: US/09/439,261
; CURRENT FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 08/833,610
; PRIOR FILING DATE: 1997-04-11
```



```
; PRIOR APPLICATION NUMBER: PCT/US98/07422
; PRIOR FILING DATE: 1998-04-10
; PRIOR APPLICATION NUMBER: US 09/227,613
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-439-261-35

Alignment Scores:
Pred. No.: 81.7 Length: 990
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-439-261-35 (1-990)

QY 72 ArgGluAlaAalaArgLeuGlyAla 79
Db 832 AGAGAGGCAGCAAGCTGGGGGCC 809

RESULT 12
US-09-227-613-34/c
; Sequence 34, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
; APPLICANT: MUKERJI, Pradip
; APPLICANT: LEONARD, Amanda E.
; APPLICANT: HUANG, Yung-Sheng
; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF
; FILE REFERENCE: 6295-US.P1
; CURRENT APPLICATION NUMBER: US/09/227,613A
; CURRENT FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 08/833,610
; PRIOR FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-227-613-34

Alignment Scores:
Pred. No.: 81.7 Length: 990
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-227-613-34 (1-990)

QY 72 ArgGluAlaAalaArgLeuGlyAla 79
Db 832 AGAGAGGCAGCAAGCTGGGGGCC 809

RESULT 13
US-09-439-261-1/c
; Sequence 1, Application US/09439261
; Patent No. 6428990
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda E.
; APPLICANT: Huang, Yung-Sheng
; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF
; FILE REFERENCE: 6295-US.P2
; CURRENT APPLICATION NUMBER: US/09/439,261
; CURRENT FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: US 08/833,610
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: PCT/US98/07422
; PRIOR FILING DATE: 1998-04-10
; PRIOR APPLICATION NUMBER: US 09/227,613
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-439-261-1

Alignment Scores:
Pred. No.: 108 Length: 1335
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-439-261-1 (1-1335)

QY 72 ArgGluAlaAalaArgLeuGlyAla 79
Db 832 AGAGAGGCAGCAAGCTGGGGGCC 809

RESULT 14
US-09-227-613-1/c
; Sequence 1, Application US/09227613A
; Patent No. 6432684
; GENERAL INFORMATION:
; APPLICANT: MUKERJI, Pradip
; APPLICANT: LEONARD, Amanda E.
; APPLICANT: HUANG, Yung-Sheng
; TITLE OF INVENTION: HUMAN DESATURASE GENE AND USES THEREOF
; FILE REFERENCE: 6295-US.P1
; CURRENT APPLICATION NUMBER: US/09/227,613A
; CURRENT FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 08/833,610
; PRIOR FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-227-613-1

Alignment Scores:
Pred. No.: 108 Length: 1335
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-227-613-1 (1-1335)

QY 72 ArgGluAlaAalaArgLeuGlyAla 79
Db 832 AGAGAGGCAGCAAGCTGGGGGCC 809

RESULT 15
US-09-039-046-1/c
; Sequence 1, Application US/09039046
; Patent No. 6331660
; GENERAL INFORMATION:
; APPLICANT: CHOMET, PAUL S.
; APPLICANT: FREY, MONIKA
```

```

; APPLICANT: GIERL, ALFONS
; TITLE OF INVENTION: MAIZE DIMBOA BIOSYNTHESIS GENES
; FILE REFERENCE: DKEM:131
; CURRENT APPLICATION NUMBER: US/09/039,046
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (49)..(1089)
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
US-09-039-046-1

```

```

Alignment Scores:
Pred. No.:      121      Length:      1508
Score:           8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      2.45%      Indels: 0
DB:                4      Gaps: 0

```

US-09-357-675C-21 (1-327) x US-09-039-046-1 (1-1508)

```

QY  302 ArgGlnLeuArgArgHisLeuPro 309
    |||
Db  585 CGACAGCTCCGGCGTCACCTCGCT 562

```

Search completed: December 8, 2002, 12:02:50  
Job time : 56 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 8, 2002, 10:39:26 ; Search time 58 Seconds  
(without alignments)  
2198.414 Million cell updates/sec

Title: us-09-357-675C-21

Perfect score: 327  
Sequence: 1 MLGFTTTPPHRFLSLCPGL.....LPVFQRRPDLVGNLGHPLS 327

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 350425 seqs, 194966369 residues

Word size: 1

Total number of hits satisfying chosen parameters: 699917

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09357675/runat\_03122002\_142237\_20448/app.query.fasta\_1.519  
-DB=PublisedApplications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09357675 @CGN\_1\_1\_21 @runat\_03122002\_142237\_20448  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAEXT=60 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database :

- Published Applications\_NA:\*
- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
  - 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
  - 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
  - 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
  - 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	253	77.4	1382	10	US-09-925-300-692
c 2	63	19.3	377	9	Sequence 692, App
c 3	63	19.3	377	9	Sequence 1096, Ap
4	13	4.0	604	10	Sequence 1096, Ap
					Sequence 929, App

5	8	2.4	56	10	US-09-866-514-5	Sequence 5, Appli
c 6	8	2.4	115	10	US-09-870-433-18	Sequence 18, Appl
c 7	8	2.4	115	10	US-09-893-191-3	Sequence 3, Appli
8	8	2.4	120	10	US-09-969-373-1197	Sequence 1197, Ap
9	8	2.4	198	10	US-09-969-373-1198	Sequence 1198, Ap
c 10	8	2.4	250	10	US-09-764-847-1703	Sequence 1703, Ap
c 11	8	2.4	260	10	US-09-878-574-10410	Sequence 10410, A
c 12	8	2.4	334	10	US-09-867-550-1983	Sequence 1983, Ap
c 13	8	2.4	351	10	US-09-870-433-17	Sequence 17, Appl
c 14	8	2.4	351	10	US-09-893-191-4	Sequence 4, Appli
c 15	8	2.4	380	10	US-09-917-800A-1047	Sequence 1047, Ap
c 16	8	2.4	762	10	US-09-910-943-678	Sequence 678, App
c 17	8	2.4	1042	10	US-09-844-493-12	Sequence 12, Appl
c 18	8	2.4	1042	10	US-09-844-501-12	Sequence 12, Appl
c 19	8	2.4	1042	10	US-09-844-265-12	Sequence 12, Appl
c 20	8	2.4	1107	9	US-09-764-868-539	Sequence 539, App
21	8	2.4	1896	10	US-09-822-849A-255	Sequence 255, App
22	8	2.4	1972	10	US-09-822-849A-485	Sequence 485, App
23	8	2.4	2210	12	US-10-052-586-287	Sequence 287, App
24	8	2.4	2212	9	US-09-764-868-166	Sequence 166, App
c 25	8	2.4	2217	9	US-09-764-868-166	Sequence 166, App
c 26	8	2.4	2324	10	US-09-917-800A-1623	Sequence 1623, Ap
c 27	8	2.4	3139	10	US-09-960-428-14	Sequence 14, Appl
c 28	8	2.4	3300	10	US-09-379-931-6	Sequence 6, Appli
29	8	2.4	3350	10	US-09-880-107-2264	Sequence 2264, Ap
c 30	8	2.4	5011	10	US-09-764-847-1562	Sequence 1562, Ap
c 31	8	2.4	6284	10	US-09-764-877-3643	Sequence 3643, Ap
c 32	8	2.4	33239	10	US-09-814-950-3	Sequence 3, Appli
c 33	8	2.4	45839	12	US-10-025-187-3	Sequence 3, Appli
c 34	8	2.4	58837	10	US-09-982-091A-5	Sequence 5, Appli
c 35	8	2.4	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c 36	7	2.1	27	10	US-09-374-671-75	Sequence 75, Appl
37	7	2.1	43	10	US-09-741-849-3	Sequence 3, Appli
38	7	2.1	48	10	US-09-865-578-9	Sequence 9, Appli
39	7	2.1	174	10	US-09-783-580-5923	Sequence 5923, Ap
c 40	7	2.1	174	10	US-09-974-300-4734	Sequence 4734, Ap
41	7	2.1	181	10	US-09-865-578-10	Sequence 10, Appl
c 42	7	2.1	202	10	US-09-867-701-8310	Sequence 8310, Ap
c 43	7	2.1	232	10	US-09-864-761-22886	Sequence 22886, A
c 44	7	2.1	235	10	US-09-923-876-3099	Sequence 3099, Ap
c 45	7	2.1	239	10	US-09-777-564-242	Sequence 242, App

ALIGNMENTS

RESULT 1  
US-09-925-300-692  
; Sequence 692, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 692  
; LENGTH: 1382  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-300-692

Alignment Scores:  
Pred. No.: 1,11e-243 Length: 1382  
Score: 253.00 Matches: 326  
Percent Similarity: 99.69% Conservative: 0  
Best Local Similarity: 99.69% Mismatches: 1



Alignment Scores:		
Pred. No.:	3.82	56
Score:	8.00	Length:
Percent Similarity:	100.00%	Matches:
Best Local Similarity:	100.00%	Conservative:
Query Match:	2.45%	Mismatches:
		Indels:

DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-866-514-5 (1-56)

QY 286 SerGluGlyProGlyLeuCysLeu 293

DB 30 TCAGAGGCGCTGGATTATGCCTC 53

#### RESULT 6

US-09-870-433-18/C  
; Sequence 18, Application US/09870433  
; Patent No. US20010034048A1  
; GENERAL INFORMATION:  
; APPLICANT: Nurith KURN  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LINEAR  
; TITLE OF INVENTION: ISOTHERMAL AMPLIFICATION OF POLYNUCLEOTIDE SEQUENCES  
; FILE REFERENCE: 492692000101  
; CURRENT APPLICATION NUMBER: US/09/870,433  
; CURRENT FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 09/660,877  
; PRIOR FILING DATE: 2000-09-13  
; PRIOR APPLICATION NUMBER: 60/175,780  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/153,604  
; PRIOR FILING DATE: 1999-09-13  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 115  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: IA013  
US-09-870-433-18

Alignment Scores:  
Pred. No.: 7.41 Length: 115  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-870-433-18 (1-115)

QY 189 SerProValSerThrProAlaGly 196

DB 105 TCTCCAGTGTCCACCCCTCCGGG 82

#### RESULT 7

US-09-893-191-3/C  
; Sequence 3, Application US/09893191  
; Patent No. US20020058270A1  
; GENERAL INFORMATION:  
; APPLICANT: Kurn, Nurith  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
; TITLE OF INVENTION: TRANSCRIPTION-BASED NUCLEIC ACID AMPLIFICATION  
; FILE REFERENCE: 49269-20006.00  
; CURRENT APPLICATION NUMBER: US/09/893,191  
; CURRENT FILING DATE: 2001-06-26  
; PRIOR APPLICATION NUMBER: 60/213,908  
; PRIOR FILING DATE: 2000-06-26  
; PRIOR APPLICATION NUMBER: 60/277,748  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 115  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Primer: IA013  
US-09-893-191-3

Alignment Scores:  
Pred. No.: 7.41 Length: 115  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-893-191-3 (1-115)

QY 189 SerProValSerThrProAlaGly 196

DB 105 TCTCCAGTGTCCACCCCTCCGGG 82

#### RESULT 8

US-09-969-373-1197  
; Sequence 1197, Application US/09969373  
; Patent No. US20020133852A1  
; GENERAL INFORMATION:  
; APPLICANT: Effertz, Roger J.  
; APPLICANT: Hauge, Brian M.  
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping  
; FILE REFERENCE: 38-10(52679)A  
; CURRENT APPLICATION NUMBER: US/09/969,373  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US 09/754,853  
; PRIOR FILING DATE: 2001-01-05  
; PRIOR APPLICATION NUMBER: US 09/760,427  
; PRIOR FILING DATE: 2001-01-13  
; PRIOR APPLICATION NUMBER: US 09/855,768  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 4593  
; SEQ ID NO 1197  
; LENGTH: 120  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-969-373-1197

Alignment Scores:  
Pred. No.: 7.71 Length: 120  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-969-373-1197 (1-120)

QY 9 ProHisArgPheLeuSerLeuLeu 16

DB 36 CCTCATAGATTCCTATCCCTCCTT 59

#### RESULT 9

US-09-969-373-1198  
; Sequence 1198, Application US/09969373  
; Patent No. US20020133852A1  
; GENERAL INFORMATION:  
; APPLICANT: Effertz, Roger J.  
; APPLICANT: Hauge, Brian M.  
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping  
; FILE REFERENCE: 38-10(52679)A  
; CURRENT APPLICATION NUMBER: US/09/969,373  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US 09/754,853  
; PRIOR FILING DATE: 2001-01-05  
; PRIOR APPLICATION NUMBER: US 09/760,427  
; PRIOR FILING DATE: 2001-01-13  
; PRIOR APPLICATION NUMBER: US 09/855,768  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 4593  
; SEQ ID NO 1198  
; LENGTH: 198

; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-969-373-1198

Alignment Scores:  
Pred. No.: 12-2 Length: 198  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-969-373-1198 (1-198)

QY 9 ProHisArgPheLeuSerLeuLeu 16  
DB 14 CCTCATAGATTCTTATCCCTCCTT 37

## RESULT 10

; Sequence 1703, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1703  
; LENGTH: 250  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-847-1703

Alignment Scores:  
Pred. No.: 15-2 Length: 250  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-764-847-1703 (1-250)

QY 212 SerLeuAlaLeuAlaGlnAlaGly 219  
DB 225 AGTCGCGCTCTTGCCGAGCTGGA 202

## RESULT 11

; Sequence 10410, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 10410  
; LENGTH: 260  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701103644H1  
US-09-878-574-10410

Alignment Scores:  
Pred. No.: 15-7 Length: 260  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-878-574-10410 (1-260)

QY 39 IleSerSerSerSerCysGluLeu 46  
DB 66 ATCTCTCTCTCTCTCTGTAATC 43

## RESULT 12

; Sequence 1983, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells a  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1983  
; LENGTH: 334  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (334)  
; OTHER INFORMATION: wherein n is one of a or t or c or g  
US-09-867-550-1983

Alignment Scores:  
Pred. No.: 19-8 Length: 334  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-867-550-1983 (1-334)

QY 9 ProHisArgPheLeuSerLeuLeu 16  
DB 27 CCACACCGATTCTTTCTTTGTC 4

## RESULT 13

; Sequence 17, Application US/09870433  
; Patent No. US20010034048A1  
; GENERAL INFORMATION:  
; APPLICANT: Nurith KURN  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR LINEAR  
; TITLE OF INVENTION: ISOTHERMAL AMPLIFICATION OF POLYNUCLEOTIDE SEQUENCES  
; FILE REFERENCE: 492692000101  
; CURRENT APPLICATION NUMBER: US/09/870,433  
; CURRENT FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 09/660,877  
; PRIOR FILING DATE: 2000-09-13  
; PRIOR APPLICATION NUMBER: 60/175,780  
; PRIOR FILING DATE: 2000-01-12

; PRIOR APPLICATION NUMBER: 60/153,604  
; PRIOR FILING DATE: 1999-09-13  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: PCR Primer  
US-09-870-433-17

Alignment Scores:  
Pred. No.: 20.7 Length: 351  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-870-433-17 (1-351)

QY 189 SerProValSerThrProAlagly 196

Db 105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 14

US-09-893-191-4/c

; Sequence 4, Application US/09893191  
; Patent No. US20020058270A1  
; GENERAL INFORMATION:

; APPLICANT: Kurn, Nurith

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR

; TITLE OF INVENTION: TRANSCRIPTION-BASED NUCLEIC ACID AMPLIFICATION

; FILE REFERENCE: 49269-20006.00

; CURRENT APPLICATION NUMBER: US/09/893,191

; PRIOR FILING DATE: 2001-06-26

; PRIOR APPLICATION NUMBER: 60/213,908

; PRIOR FILING DATE: 2000-06-26

; PRIOR APPLICATION NUMBER: 60/277,748

; PRIOR FILING DATE: 2001-03-21

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 351

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic Primer

US-09-893-191-4

Alignment Scores:

Pred. No.: 20.7 Length: 351  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-893-191-4 (1-351)

QY 189 SerProValSerThrProAlagly 196

Db 105 TCTCCAGTGTCCACCCCTGCCGGG 82

RESULT 15

US-09-917-800A-1047/c

; Sequence 1047, Application US/09917800A

; Patent No. US20020119462A1

; GENERAL INFORMATION:

; APPLICANT: Mendrick, Donna

; APPLICANT: Porter, Mark

; APPLICANT: Johnson, Kory

; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1047  
; LENGTH: 380  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1178850  
US-09-917-800A-1047

Alignment Scores:

Pred. No.: 22.3 Length: 380  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.45% Indels: 0  
DB: 10 Gaps: 0

US-09-357-675C-21 (1-327) x US-09-917-800A-1047 (1-380)

QY 282 ValAlaArgCysSerGluGlyPro 289

Db 194 GTGCAAGGTGTAGTGAGGACCT 171

Search completed: December 8, 2002, 12:04:09

Job time : 61 secs